THE TEACHING OF MANDARIN PROSODY: 
A SOMATICALLY-ENHANCED APPROACH 
FOR SECOND LANGUAGE LEARNERS

Felicia Zhen Zhang

A thesis submitted in fulfilment of the requirements for
the degree of

Doctor of Philosophy in the Division of Communication 
and Education

University of Canberra

© Felicia Zhang 2006

All rights reserved. This work may not be
reproduced in whole or in part, by photocopy
or other means, without the permission of the author.
ABSTRACT

For adult English speakers studying Mandarin (Modern Standard Chinese), the acquisition of the Mandarin prosody presents major difficulties. One particularly problematic aspect of the Mandarin prosodic system, and the one singled out for research here is the acquisition of tones by second language (L2) learners of Mandarin. This thesis involves a literature review and a description of an experiment conducted for the purpose of assessing the effectiveness of a new teaching method for educating students in Mandarin prosody generally, but especially with regard to "tones."

Most studies investigating the acquisition of Mandarin tones by L2 learners have treated tones as separate from other aspects of Mandarin prosody such as stress, loudness and duration. The teaching method examined in this thesis, however, takes an alternative approach. Here the acquisition of Mandarin prosody is approached as a complex dynamic that has tones as an integral part.

The aims of the study are twofold: (1) to identify the principal problems encountered by most learners in order to discover the causes of recurrent error patterns and, (2) to find out how a multi-sensory approach, which in this study was called the Somatically Enhanced Approach (SEA), might influence the acquisition of Mandarin prosody in these areas.

The experiment involved 22 adult Australian students studying Mandarin in the first three months of language training. The experimental component of the study consisted of an evaluation of two groups of students’ oral conversations. The two groups of students were divided into a control group and an experimental group. The control group was trained in a non-multi-sensory but communicative approach in 2001 and 2002. Their results are compared with those of a test group and with a group of students trained in the multi-sensory communicative approach (SEA) in 2003 and 2004. The test materials consisted of short dialogues that were likely to occur in everyday communication. Data was collected from each group, once during the first half of the first semester of study in each year.

The findings of the experiment were that the order of difficulty of the four Mandarin tones was found to be similar for both the experimental and control groups of students. However, the order
of difficulty differed from what has been reported by previous researchers. This suggests that
the input and the type of task used to collect data might exert a significant influence on the
learning of tones. In other words, the performance of subjects in the dialogues suggests that in
the initial stages of learning, the major cause of errors was first language (L1) interference
rather than the physical "difficulty" of articulating particular phonemes (or any features of
Universal Grammar). Therefore, by using a multi-sensory approach (SEA) to the learning of
Mandarin, it may be possible to considerably lessen the influence of learners’ L1 from the
outset.

Finally, a number of suggestions for improving the teaching of Mandarin prosody are made and
future research directions outlined. Some salient suggestions for teaching of Mandarin prosody
that arise from the research are:

(1) To use movement and gesture in the early stages of learning to enhance students’ perception
and production of Mandarin. This approach provides students with useful memory tools for
learning both in class and in self-accessed learning;

(2) To teach Tone 3 not as a full Tone 3 but as a low level tone. This should not be done solely
through a simple verbal explanation but through a combination of movement and gesture,
provision of visual and auditory feedback and a large amount of exposure and perception
training so that Tone 3 is recognised as a low level tone rather than a full Tone 3. By so doing
confusion is reduced between the various realizations of Tone 3 during the initial learning
stages; and

(3) To caution students about the common error patterns caused by interference from their L1.
This should be supplemented with opportunities for students to observe their own production of
Mandarin and then experience how physically they can find ways of reducing the interference.

A qualitative analysis of interview and question data obtained from this research also revealed
that the extensive use of computer enhanced language learning and SEA work well together, not
only efficiently conditioning students to the phonology of Mandarin, but dramatically changing
students' strategies in learning and increasing their learning opportunities.
ACKNOWLEDGMENTS

I would like to express my profound gratitude to an inspirational scholar such as Professor Andrew Lian. Since we met, he has been a constantly source of encouragement and inspiration. I feel extremely lucky to have met a scholar who has had the courage to challenge the status quo of language teaching. Had it not been for his unique insights and depth of thinking, I would not have so bravely embarked on this experiment with a new teaching methodology in Mandarin teaching.

I am also grateful to Associate Professor Michael Sawer for being my secondary supervisor and Professor Michael Wagner for his constant encouragement and for being on my supervisory panel. Through doing this project, I have managed to learn a great deal about speech recognition and acoustic phonetics from Professor Michael Wagner. For this broadening of my knowledge, I am truly grateful.

My sincere thanks also go to Mr. Chris McMahon, who is also an inspiring thinker and my husband. Without his unfailing support and constant questioning of various parts of my thesis, I believe I would not have been able to complete this task in a satisfactory manner. I would also like to thank my beautiful daughters XinXin and PeiPei for their laughter and cuddles and their own unique ways of learning which have impacted on the way this thesis was conceptualised a great deal.

I would also like to thank my mother and sister for their support and in particular for starting me on the journey of learning English in a sports stadium in China so many years ago. Without my mother’s insistence and bravery, this thesis would not be written today.

I also like to thank the many friends and colleagues I have met or worked with since commencing this project. Without their cheerful support, I would not be able to stay sane for so long.
# TABLE OF CONTENTS

ABSTRACT................................................................................................................................................... ii
ACKNOWLEDGMENTS........................................................................................................................... v
LIST OF TABLES........................................................................................................................................... xiii
LIST OF FIGURES......................................................................................................................................... xv
GLOSSARY................................................................................................................................................xvi
Chapter 1 Introduction ................................................................................................................................ 1
  1.1 The study .............................................................................................................................................. 1
  1.2 Problem and Rationale ..................................................................................................................... 1
  1.3 Objectives .......................................................................................................................................... 4
  1.4 Chapter Outline .................................................................................................................................. 5
Chapter 2 Mandarin Phonological System .............................................................................................. 7
  2.1 The phonology of Mandarin ............................................................................................................ 7
      2.1.1 Initials ......................................................................................................................................... 7
      2.1.2 Finals ......................................................................................................................................... 10
      2.1.3 Tones ......................................................................................................................................... 11
      2.1.4 The Notational system of tones ............................................................................................. 11
  2.2 Mandarin prosodic system .............................................................................................................. 14
      2.2.1 Fundamental frequency, duration, amplitude and tone......................................................... 15
      2.2.2 Word and phrase features: stress ......................................................................................... 16
      2.2.3 Tone and Intonation ............................................................................................................... 17
  2.3 Summary ............................................................................................................................................ 21
Chapter 3 Review of Literature ............................................................................................................ 22
  3.1 Tone Acquisition by Mandarin-speaking children as a L1 ............................................................ 22
      3.1.1 Studies on the acquisition of Mandarin tones .................................................................. 22
      3.1.2 Relationship between perception and production ............................................................. 32
      3.1.3 Acquisition of tones before segments ............................................................................. 36
  3.2 Tone acquisition of Mandarin in L2 environments ..................................................................... 39
      3.2.1 Research on the tone acquisition of Mandarin in L2 environments .................................... 39
          3.2.1.1 Descriptive studies of tonal errors ........................................................................ 39
          3.2.1.2 What changes occur in the voices of English speakers of Mandarin when speaking Mandarin and English? ......................................................... 40
          3.2.1.3 Causes of tonal errors ............................................................................................... 40
          3.2.1.4 Order of difficulty ..................................................................................................... 46
          3.2.1.5 Research into the pedagogy of teaching tones ............................................................ 51
      3.2.2 Pedagogic measures to train students’ awareness of tone registers .................................... 55
      3.2.3 Review of current research on Mandarin Prosody ............................................................... 57
3.2.4 Summary of research on tone acquisition by Mandarin speaking children as L1 and by students learning Mandarin as L2

3.3 Child-Adult Differences in Language Acquisition

3.3.1 The Critical Period Hypothesis

3.3.1.1 What is a critical period?

3.3.1.2 Critical Periods in Language Acquisition

3.3.1.3 Evidence of perceptual sensitivities not being lost

3.3.1.4 Plasticity of the brain

3.3.1.5 The effect of training after the optimum period

3.3.2 The Less-Is-More Principle

3.3.3 Cognitive Explanations

3.3.3.1 Left brain or right brain

3.3.3.2 Production Processes in SLA

3.3.3.3 Percept-behaviour Expressway (argument for the relationship between perception and production)

3.3.4 Social-Psychological Factors

3.3.4.1 Individual Factors of Acculturation

3.3.5 Pedagogical Environment

3.3.5.1 Affective factors

3.3.5.2 Role of the teacher

3.3.5.3 A paradigm shift: from language learning through prescription to language learning through discovery

3.3.5.3.1 Individualization

3.3.5.3.2 Activity Space

3.4 Pronunciation Pedagogy

3.4.1 What do we teach? Segments or suprasegmentals? Inseparable!

3.4.2 Teaching segmentals

3.4.2.1 Tonal contrast pairs

3.4.2.2 Minimal pairs and the language laboratory

3.5 Teaching Mandarin tones

3.5.1 Traditional pedagogy

3.5.1.1 Error Correction

3.6 The shift to focus on the suprasegmental in teaching pronunciation, is it warranted?

3.6.1 Evidence from L1 acquisition

3.6.1.1 Sensitivity to Prosodic markers in the Input

3.6.1.2 Evidence that prosodic organization is actually used in infant’s speech processing

3.7 Rhythm

3.7.1 What else do infants pick up from their interaction with their carers?

3.7.2 Rhythmicity as a Source of Prediction

3.8 Body movements
### 3.8.1 Review of therapeutic use of body movement

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.8.1.1 Verbotonalism</td>
<td>110</td>
</tr>
<tr>
<td>3.8.1.2 Speech-movement therapy based on sense physiology</td>
<td>114</td>
</tr>
</tbody>
</table>

### 3.9 Learner-centred classroom approaches

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.9.1 Learning through multi-modalities</td>
<td>123</td>
</tr>
</tbody>
</table>

### 3.10 Computer-Assisted Pronunciation Training

### 3.11 Essential features of a Somatically Enhanced Approach in the creation of a Mandarin language learning environment

### 3.12 Summary

### Chapter 4 Toward a Somatically Enhanced language learning environment for teaching Mandarin tones and prosody

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.1 Introduction</td>
<td>141</td>
</tr>
<tr>
<td>4.2 Phonological Model</td>
<td>145</td>
</tr>
<tr>
<td>4.3 Classroom methodology</td>
<td>149</td>
</tr>
<tr>
<td>4.3.1 Sensitization Session</td>
<td>150</td>
</tr>
<tr>
<td>4.3.1.1 Step 1: Relaxation</td>
<td>150</td>
</tr>
<tr>
<td>4.3.1.2 Step 2: Humming</td>
<td>153</td>
</tr>
<tr>
<td>4.3.1.3 Step 3: Clapping to the rhythm of the sentences</td>
<td>155</td>
</tr>
<tr>
<td>4.3.1.4 Step 4: Incorporation of movement and gesture</td>
<td>156</td>
</tr>
<tr>
<td>4.3.1.4.1 Tension</td>
<td>157</td>
</tr>
<tr>
<td>4.3.1.5 Step 5: Mouthing the words</td>
<td>162</td>
</tr>
<tr>
<td>4.3.1.6 Step 6-7: Adding words to the intonation patterns</td>
<td>163</td>
</tr>
<tr>
<td>4.3.2 Repetition exercises</td>
<td>163</td>
</tr>
<tr>
<td>4.3.2.1 Step 8 Checking for meaning</td>
<td>165</td>
</tr>
<tr>
<td>4.3.3 Exposure to authentic materials</td>
<td>165</td>
</tr>
<tr>
<td>4.3.4 Choice of teaching materials</td>
<td>167</td>
</tr>
<tr>
<td>4.3.5 Selection of Vocabulary</td>
<td>170</td>
</tr>
<tr>
<td>4.3.6 Use the text only as a last resort</td>
<td>171</td>
</tr>
<tr>
<td>4.3.7 Role of the teacher</td>
<td>172</td>
</tr>
<tr>
<td>4.3.7.1 Supervision of the sensitization session</td>
<td>172</td>
</tr>
<tr>
<td>4.3.7.2 Monitoring of learner performances</td>
<td>173</td>
</tr>
<tr>
<td>4.3.7.3 Establishing links to real face-to-face communication</td>
<td>173</td>
</tr>
<tr>
<td>4.3.7.4 Teaching learners how to learn</td>
<td>174</td>
</tr>
<tr>
<td>4.4 Computer enhanced learning support</td>
<td>175</td>
</tr>
<tr>
<td>4.4.1 Teaching materials</td>
<td>175</td>
</tr>
<tr>
<td>4.4.2 Function of Sptool</td>
<td>176</td>
</tr>
<tr>
<td>4.4.3 Role of Sptool</td>
<td>180</td>
</tr>
<tr>
<td>4.5 Feedback provision</td>
<td>183</td>
</tr>
<tr>
<td>4.6 Individualization of the curriculum</td>
<td>183</td>
</tr>
</tbody>
</table>
6.2.3 The effects of Somatically Enhanced Approach on the Fundamental Frequency (F₀) of the subjects ................................................................. 219

6.3 Results of the perceptual experiment (Experiment two) ...................................................... 225

6.4 Error analysis of the Tonal errors ..................................................................................... 226
   6.4.1 Summary: control group ........................................................................................ 226
   6.4.2 Summary: experimental group .............................................................................. 229
   6.4.3 Order of difficulty of tones .................................................................................... 231

6.5 Quality of data produced by both experimental and control groups ............................... 236
   6.5.1 Quality of oral performances from the control group ........................................... 237
   6.5.2 Quality of the oral performances of the experimental group ............................... 237
   6.5.3 Summary on the quality and quantity of the oral performances from both experimental and control groups ................................................. 240
   6.5.4 Description of Tone errors ..................................................................................... 241
      6.5.4.1 Tone one – High Level ................................................................................. 241
      6.5.4.2 Tone two – Rising ......................................................................................... 244
      6.5.4.3 Tone three - Low Dipping ............................................................................ 247
      6.5.4.4 Tone four – High Falling .............................................................................. 249
      6.5.4.5 Tone five - The Neutral Tone ....................................................................... 252

6.6 Types of Errors ............................................................................................................ 253
   6.6.1 Interference Errors ................................................................................................. 253
   6.6.2 Induced errors ......................................................................................................... 256

6.7 Summary .................................................................................................................... 257

Chapter 7 Qualitative results for both control and experimental groups of students ............ 260

7.1 Control group .............................................................................................................. 261
   7.1.1 Results of end of course questionnaires ............................................................... 261
   7.1.2 Interview data ....................................................................................................... 261
      7.1.2.1 The control group’s use of cognitive strategies in language learning .......... 262
         7.1.2.1.1 Direct strategy use of the control group .............................................. 262
         7.1.2.1.2 Indirect strategy use of the control group ........................................... 265
   7.1.2.2 Need to work on the tones ............................................................................. 267
   7.1.2.3 Reaction to not using romanization ............................................................... 267

7.2 Experimental group ..................................................................................................... 269
   7.2.1 Results of the computer questionnaires ............................................................... 269
   7.2.2 Interview data ....................................................................................................... 269
      7.2.2.1 Direct strategy use of the experimental group ............................................. 269
         7.2.2.1.1 Variety of memory-enhancing strategies used ................................ 271
         7.2.2.1.2 How did the students practise the language? ...................................... 272
   7.2.2.2 Use of indirect strategies by the experimental group ...................................... 277
   7.2.2.3 A student’s reaction to Sptool ....................................................................... 279
   7.2.2.4 Students’ reaction to not using Pinyin............................................................ 279
7.3 Summary ......................................................................................................................... 281

Chapter 8 Conclusions ........................................................................................................... 283

8.1 Information obtained about the process of prosody acquisition by English speaking
Australian students learning Mandarin as a second language ............................................ 283

8.1.1 Order of difficulty with the four Mandarin tones ....................................................... 284
8.1.2 Common tone error patterns and causes among learners ......................................... 286
8.1.3 Likely causes for the detected errors ...................................................................... 288

8.2 Discussion of principles ................................................................................................. 292

8.2.1 Contribution of gesture and movement to the teaching of Mandarin prosody .... 292
8.2.2 Language learning needs to be carried out in its social context and access to the TL
discourse community is important .................................................................................. 294
8.2.3 The role of CD-ROMs ............................................................................................... 295

8.3 Implications for course design ....................................................................................... 295

8.4 Contribution of this thesis to the state of knowledge ..................................................... 298

8.5 Unresolved questions ..................................................................................................... 300

8.5.1 Role of romanization in Mandarin language teaching .............................................. 300
8.5.2 The use of Sptool ...................................................................................................... 301
8.5.3 Is this methodology similar to that of the Total Physical Response (TPR)? ............. 302
8.5.4 Does metalanguage aid learning? ............................................................................ 303

8.6 Future research directions ............................................................................................. 305

8.6.1 Some new questions to be asked .............................................................................. 305

8.6.1.1 Do L2 learners of Mandarin segment Mandarin as L1 children do? .................. 305
8.6.1.2 Does this study support the critical age theory of language acquisition? ....... 305
8.6.1.3 Is there any evidence to support Krashen’s “input hypothesis”? ..................... 307
8.6.1.4 Is SLA like L1 Acquisition? ............................................................................ 308
8.6.1.5 Future research ............................................................................................... 308

8.7 Concluding Remarks ...................................................................................................... 310

Appendices ...................................................................................................................... 313

Appendix One: Sample teaching material ........................................................................... 313
Appendix Two: Content page of the course ......................................................................... 318
Appendix Three: Teaching materials covered in the first six weeks ............................... 319
Appendix Four: Guiding Questions for Interview ............................................................... 323
Appendix Six: Transcription of a conversation between Student 1 and a Thai student from
experimental group ............................................................................................................. 325
Appendix Seven: Transcription of a conversation between 2 students in the control group .... 328
LIST OF TABLES

Table 2.1. Mandarin Initials (consonants)...................................................................................... 8
Table 2.2. Correspondence between IPA and Pinyin Symbols...................................................... 9
Table 2.3. Mandarin Finals (vowels) ............................................................................................10
Table 2.4. Figure of Chinese Tones ............................................................................................. . 12
Table 6.1 The group averages of speakers’ mean F0, mean F0 SD, semitone, mean minimum F0, mean maximum F0 and mean F0 range for 5 groups of male speakers ......................... 206
Table 6.2 T-tests for group average mean F0, average mean F0 SD and semitone, minimum F0, maximum F0, F0 range, of the male speakers in the experimental group, for Mandarin and English languages (in Hz) ................................................................................ 208
Table 6.3 T-tests for group average mean F0, average mean F0 SD and semitone, minimum F0, maximum F0, F0 range, of the male speakers in the control group, for Mandarin and English languages (in Hz) ........................................................................... 210
Table 6.4 One-way ANOVA test comparing group average mean F0, average mean F0 SD and semitone, minimum F0, maximum F0, F0 range, of the male speakers in all groups in Mandarin (in Hz) .............................................................................. 211
Table 6.5 Post Hoc tests for one way ANOVA for the dependent variable average mean F0 for Mandarin between experimental, control and NS groups of male subjects. .................. 212
Table 6.6 Post Hoc tests for one way ANOVA for the dependent variable average mean F0 SD for Mandarin between experimental, control and NS groups of male subjects. .......... 212
Table 6.7 Post Hoc tests for one way ANOVA for the dependent variable semitone for Mandarin between experimental, control and NS groups of male students. ................... 213
Table 6.8 The group averages of speakers’ mean F0, minimum F0, maximum F0 and F0 range for 5 groups of female speakers............................................................................................ 214
Table 6.9 T-tests for group average mean F0, average mean F0 SD and semitone, minimum F0, maximum F0, F0 range, of the female speakers in the experimental group, for Mandarin and English languages (in Hz) ................................................................................ 215
Table 6.10 T-tests for group average mean F0, average mean F0 SD and semitone, minimum F0, maximum F0, F0 range, of the female speakers in the control group, for Mandarin and English languages (in Hz) ........................................................................... 216
Table 6.11 One-way ANOVA test comparing group average mean F0, average mean F0 SD and semitone, minimum F0, maximum F0, F0 range, of the female speakers in all groups in Mandarin (in Hz) .............................................................................. 217
Table 6.12 Post Hoc tests for one way ANOVA for the dependent variable minimum F0 for Mandarin between experimental, control and NS groups of male subjects. .................. 218
Table 6.13 Post Hoc tests for one way ANOVA for the dependent variable maximum F0 for Mandarin between experimental, control and NS groups of female subjects. ......... 219
Table 6.14 T-tests for group average mean F0, average mean F0 SD and minimum F0, maximum F0, F0 range, of the male speakers in Chen’s study (in Hz) (n=2) .......................... 224
Table 6.15 T-tests for group average mean F0, average mean F0 SD and minimum F0, maximum F0, F0 range, of the female speakers in Chen’s study (in Hz) (n=2) .......................... 224
Table 6.16 T-test analysis for dependent samples illustrating the relation between experimental and control groups and the dependent sample (mean perceptual rating scores given by NS judges) . .......................................................... 225
Table 6.17 Distribution of different tones and tone errors in control group’s oral production. 227
Table 6.18 Error distribution per tone for each subject in the control group............................. 228
Table 6.19 Distribution of different tones and tonal errors in experimental group’s oral production................................................................. 229
Table 6.20 Error distribution per tone for the experimental group............................................ 230
Table 6.21 Order of difficulty for Mandarin Tones by different researchers (including this study) in the field of teaching Mandarin as a FL (CFL) and L1 acquisition by Mandarin children (L1 of Mandarin)................................................................. 232
Table 6.22 Quality of data produced by the control group.......................................................... 237
Table 6.23 Quality of data produced by the experimental group............................................... 238
Table 6.24 Production of Tone 1 from 5 students in the experimental group ........................... 242
Table 6.25 Production of Tone 1 from 2 NSs of Mandarin from the NS group’s data used in the study................................................................. 242
Table 6.26. Four students’ production of Tone 1 from the control group ................................. 244
Table 6.27 Production of Tone 2 from 2 NSs of Mandarin........................................................ 245
Table 6.28 Students’ production of Tone 2 from the experimental group................................ 246
Table 6.29 Students’ production of Tone 2 from the control group......................................... 246
Table 6.30 Students’ production of Tone 4 from the experimental group................................. 250
Table 6.31 Students’ production of Tone 4 from the control group.......................................... 251
Table 6.32 NSs’ production of Tone 4.................................................................................... 251
Table 7.1 Direct (performance) strategy use of four students in the control group................... 262
Table 7.2 Indirect strategy use of four students in the control group......................................... 262
Table 7.3 Direct (performance) strategy use of five students in the experimental group........... 269
Table 7.4 Indirect strategy use of five students in the experimental group. ............................... 277
## LIST OF FIGURES

<table>
<thead>
<tr>
<th>Number</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Figure 3.1 The realization of Tone 1 by Mandarin and English speakers</td>
<td>41</td>
</tr>
<tr>
<td>Figure 3.2 The realization of Tone 2 by Mandarin and English speakers</td>
<td>42</td>
</tr>
<tr>
<td>Figure 3.3 The realization of Tone 3 (half-third) by Mandarin and English speakers</td>
<td>42</td>
</tr>
<tr>
<td>Figure 3.4 The realization of Tone 3 by Mandarin and English speakers</td>
<td>42</td>
</tr>
<tr>
<td>Figure 3.5 The realization of “nǐ hǎo” by L2 speaker of Mandarin in normal and happy moods</td>
<td>45</td>
</tr>
<tr>
<td>Figure 4.1(a) The pitch curve of the sentence “Nǐ xǐhuān bù xǐhuān chī jiāozi?” in which “nǐ” is not stressed.</td>
<td>168</td>
</tr>
<tr>
<td>Figure 4.1(b) The pitch curve of the sentence “Nǐ xǐhuān bù xǐhuān chī jiāozi?” in which “nǐ” is stressed.</td>
<td>168</td>
</tr>
<tr>
<td>Figure 4.2 Sample page of the teaching material.</td>
<td>170</td>
</tr>
<tr>
<td>Figure 4.3 A screenshot of the speech analysis tool</td>
<td>178</td>
</tr>
<tr>
<td>Figure 4.4 The “Option” page of the Sptool</td>
<td>179</td>
</tr>
<tr>
<td>Figure 4.5 Picture of the Sptool showing the sample sentence: Hello, what’s your honourable surname?</td>
<td>180</td>
</tr>
<tr>
<td>Figure 4.6 Pitch curve of “wǒ shēngyú yǐjiǔ wǔ líng nián bā yuè èr shí sān hào” (我生于一九五零年八月二十号。)</td>
<td>182</td>
</tr>
<tr>
<td>Figure 6.1 NS judges’ ratings on each subject’s Mandarin oral performance in the control and experimental groups</td>
<td>225</td>
</tr>
<tr>
<td>Figure 6.2 Distribution of different tones and tone errors of control group by each subject</td>
<td>227</td>
</tr>
<tr>
<td>Figure 6.3 Error distribution per tone for the control group</td>
<td>228</td>
</tr>
<tr>
<td>Figure 6.4 Distribution of different tones and tonal errors of experimental group</td>
<td>230</td>
</tr>
<tr>
<td>Figure 6.5 Error distribution per tone for the experimental group</td>
<td>231</td>
</tr>
<tr>
<td>Figure 6.6 Mispronunciation of “wǒ zhùzài Braddon.”</td>
<td>235</td>
</tr>
<tr>
<td>Figure 6.7 Mispronunciation of “wǒ méiyǒu jiějie.”</td>
<td>236</td>
</tr>
<tr>
<td>Figure 6.8 Comparison of learner’s pitch contour (top) and the teacher’s model (bottom) of the sentence “wǒ yòu yíge jiějie, jiào Helen, tā sǎnshí suí le, tā jièhunle, yǒu yíge háizi, tā huáiyúnle”</td>
<td>239</td>
</tr>
<tr>
<td>Figure 6.9 Misproduction of “Tā wǔshíyīsuìle”</td>
<td>243</td>
</tr>
<tr>
<td>Figure 6.10. Mispronunciation of “Cóng” in “Colin cóng England lái de”</td>
<td>247</td>
</tr>
<tr>
<td>Figure 6.11 Mispronunciation of “wō” in “wō jiào X”</td>
<td>248</td>
</tr>
<tr>
<td>Figure 6.12 Mispronunciation of “wōhēn” in “rènshi nǐ, wō hěn gāoxìng”</td>
<td>249</td>
</tr>
<tr>
<td>Figure 6.13 Mispronunciation of “tā duō dà le?”</td>
<td>252</td>
</tr>
</tbody>
</table>
GLOSSARY

Words, definitions and abbreviations in alphabetic order:

WORDS AND DEFINATIONS:

**Allotone**: Tone 3 as it possesses two allotones: one dipping and one rising. In other words, though Tone 3 occurs in these two forms, the variation does not affect the tone’s functional identity in the language, i.e. no change in meaning is involved. (Cheng, 1973; Hatch & Lazaraton, 1991; Hatch, 1983; Hombert, 1975, 1989b:159)

**Downtrend**: A combination of downstep and declination. Downstep and declination are important aspects of intonation. Downstep refers to the phenomenon that a high (H) pitch target has lower F₀ height after a low (L) pitch target; while declination refers to the tendency for F₀ to gradually decline over the course of an utterance. A broad term “downtrend” is used to describe the combined effects of the two.

**Dyspraxia**: Developmental dyspraxia is an impairment or immaturity of the organization of movement. It is an immaturity in the way that the brain processes information, which results in messages not being properly or fully transmitted. The term dyspraxia comes from the word praxis, which means “doing, acting”. Dyspraxia affects the planning of what to do and how to do it. It is associated with problems of perception, language and thought. Retrieved on 17th of July, 2005, from http://www.dyspraxiafoundation.org.uk/dyspraxia-information/whatis.html

**Finals**: Vowels in Mandarin: The final is the part of the syllable excluding the initial. There are thirty-seven finals in Mandarin.

**Fundamental frequency**: A term which is derived from the study of the physics of sound, and used in acoustic phonetics. It refers to the lowest frequency component in a complex sound way (other components being known as the “harmonics”). Frequency refers to the number of complete cycles (opening and closing movements) of vocal cord vibration in the unit of time
The “fundamental”, or F₀ (“f nought”), is of particular importance in studies of intonation, where it displays a reasonably close correspondence with the pitch movements involved. It is measured in hertz (Hz), a term which has replaced the older “cycles per second”. (Crystal, 1997)

**Headturn preference procedure:** This procedure was originally developed by Fernald (1985) for her investigations of infants’ listening preferences for infant- versus adults-directed speech. She used a three-sided testing booth that was open on the fourth side. White curtains were hung between the ceiling and the top three sides of the booth in order to block the infant’s view of the rest of the room. Loudspeakers were mounted into the walls of the two panels at about the level of the infant’s head. A small red light located on each side of the panel in the vicinity of the loudspeaker. The centre panel, which the infant faced, had a small green light mounted at infant’s eye level. Directly below this light there was a 5 centimetre hole cut into the panel for the lens of a video camera. A chair was situated in the centre of the booth approximately even with the two side lights and facing the centre panel. The caregiver sat on this chair and held the infant on her lap. In her use of the headturn preference procedure, Fernald’s primary dependent measure was the direction of the infant’s first head trial. Experimental sessions began with a series of training trials and infants were familiarized with four different 8-second speech samples, available on the two sides of the booth. To start a training trial, the green light on the centre panes was flashed to draw the infant’s attention. When a judge, viewing a video monitor in another room, decided that the infant was looking straight ahead, the green light was turned off. The judge then signals another experimenter to turn on one of the two red side lights. The choice of whether the left or right side light was flashed on a trial followed a training order that was set up for a given subject. The experimenter was responsible for turning on the tape recorder that played the type of speech sample appropriate to that side (e.g., infant-directed...
speech samples on the left and adult-directed samples on the right). The assignment of sample
types to sides was counterbalanced across subjects. When a given sample ended, the red light
was extinguished and the green centre light was flashed until the infant’s gaze was centred.
Then the next trial began. If the infant did not spontaneously respond to the flashing light within
the first few seconds of a training trial, the caregiver was asked to turn the infant in that
direction. After four training, the caregiver was instructed to keep the infant centred and not to
turn the infant during the remaining trials. Both the caregiver and the judge wore headphones
and listened to recorded music to mask the speech samples presented during the experiment.
During the test phase, the presentation of a speech sample was made contingent upon a 30°
headturn by the infant. As in the training trials, the green light was used to attract the infant’s
gaze to midline. After it was extinguished, the first 30° headturn to the left or right resulted in
the presentation of a speech sample appropriate to that side accompanied by the blinking red
light. Each sample was played to its completion regardless of whether the infant looked away
before it finished. Infants had to complete at least 15 test trials to be included in the study.
Subjects were scored as to the number of trials that they turned to a given side. Fernald (1985)
found that infants turned significantly more toward the side with the infant-directed speech
samples.

**Initials:** The initial represents the consonantal beginning of a syllable. Since Mandarin does not
have consonant clusters (sequence of consonants), the consonantal beginning of a syllable can
only be a single consonant. There are, however, Mandarin syllables that do not have any initial
consonants. For those syllables the tradition is to describe their initials as “zero”.


**Kinemorphae:** The term kinemorphae or movement Gestalt is used here in the meaning of the Japanese term Kata. It denotes a wide-ranging class of dynamic cultural transmissions that comprises elements such as dance, but also gymnastics, martial and marital arts, and juggling, etc., whose common denominator is that they involve complex movement patterns of the body. Retrieved on 15th of November, 2004 from http://www.uni-ulm.de/uni/intgruppen/memosys/des24n.htm#fnB531

**Mean F₀ SD:** Standard deviation of the mean fundamental frequency is the acoustic measure which measures the range of pitch (auditory) of a person’s voice level. In this thesis, it is not a measure of the highest or the lowest point of a person’s voice level.

**Pitch:** The attribute of auditory sensation in terms of which a sound may be ordered on a scale from “low” to “high”. It is an auditory phonetic feature, corresponding to some degree with the acoustic feature of frequency, which in the study of speech is based upon the number of complete cycles of vibration of the vocal cords. Variations of pitch are more easily produced using voiced sounds, because of their regular wave-form. (Crystal, 1997)

**Pitch range:** Highest point of the fundamental frequency of a person’s voice through to the lowest point of the fundamental frequency of a person’s voice level.

**Proprioception:** The ability to sense stimuli arising within the body. Even if you are blindfolded, you know through proprioception if your arm is above your head or hanging by your side. The word “proprioception” was coined in 1906 by the English neurophysiologist Charles Sherrington who received the Nobel Prize for Physiology or Medicine in 1932 for research on the function of the neuron and study of reflex action. Retrieved on the 15th of November, 2004 from http://www.medterms.com/script/main/hp.asp

**Semitone:** It has been pointed out that the statistics of F₀-values are often not very well described by a normal distribution. If F₀ is scaled linearly (in Hz), there is, typically, some positive skewness (Traunmuller & Eriksson, 1994). In order make the results from different groups in this study more comparable, the average mean F₀ SD for the group is also reported in semitones. Traunmuller and Eriksson further points out that the liveliness of a discourse can be
observed by comparing the F₀-excursion of the utterances which is measured in semitones. In other words, the higher the SD of the average mean F₀ measured in semitones, the livelier the discourse (Traunmüller, 1994).

**Tension:** A term used in phonetic classification of speech sounds, referring to the overall muscular effort used in producing a sound. The contrasts are labelled variously, e.g. fortis v. Lenis, tense vs. lax. This contrast is viewed as particularly important in distinctive feature theories of phonology, where tense is one of the main features set up to handle variations in the manner of articulation. Tense sounds are sounds produced with a relatively strong muscular effort, involving a greater movement of the (supraglottal) vocal tract away from the position of rest (cf. fortis) and a relatively strong spread of acoustic energy. The opposite term in Jakobson and Halle’s system is lax. (Crystal, 1997)

**Tone sandhi rules:** Tone sandhi may be described as the change of tones when syllables are juxtaposed. To put it differently, a syllable has one of the tones in the language when it stands alone, but the same syllable may take on a different tone without a change in meaning when it is followed by another syllable. A tone sandhi rule in Mandarin involves the second tone, which changes into the first tone when it is preceded by either the first or the second tone and followed by any one of the four tones. Two important tone sandhi rules in Mandarin involve the third tone:

Tone sandhi rule 1: When a third tone syllable is followed by a syllable with any tone other than another third tone, the third-tone syllable changes to a low-tone syllable with the pitch contour 21. For example, mă “horse” has the third tone in isolation, but when it is followed by another syllable such as chē “vehicle”, the sequence is pronounced with following tone sequence: mā(21) chē(55).

Tone sandhi rule 2: When a third tone syllable is followed by another third tone syllable, the first one changes into a second tone. For example, gān “to chase” and guī “demon” both have third tones. When they are in sequence, gān guī “to exorcise demons”, gān is change from third tone (214) to second tone (35)
Neutral tone: if a syllable has a weak stress or is unstressed, it loses its contrastive, relative pitch and therefore does not have one of the four tones described above. In such a case, the syllable is said to have a neutral tone. According to Chao (1968:3), the pitch of the neutral tone is:

- Half-low after first tone: tā-de-his;
- Middle after second tone: hóng-de- red one;
- Half-high after third tone: wō-de- my;
- Low after fourth tone lù-de=green one.

(Neutral tone in this thesis is indicated with a number 5 after the syllable. In some books, the neutral tone is represented by 0 rather than 5) (Li & Thompson, 1989)

**VTM**: Verbo-tonal method of phonetic correction. Method of speech rehabilitation for persons with communication impairments, based upon body movements which imitate articulatory movements of speech or based upon amplification frequencies that are not perceived. (Blouin & Bergeron, 1997:49).

**ABBREVIATIONS:**

**AE**: American English

**ASL**: American Sign Language

**ASR**: Automatic Speech Recognition

**BD**: Beijing dialect on which Mandarin is based on.

**CAL**: Computer Assisted Learning

**CALL**: Computer Assisted Language Learning

**CAPT**: Computer Assisted Pronunciation Technology
**CP:** Critical Period

**ERP:** event related potential

**FD:** Fundamental Difference

**FL:** FL

**IL:** Interlanguage

**L1:** First Language

**LAD:** Language Acquisition Device

**MLU:** Mean Length of Utterances

**NS:** native speaker

**NNS:** non-native speaker

**L1:** First Language

**L2:** Second Language

**OP:** Optimum Period

**SEA:** Somatically-Enhanced Approach

**SLA:** SLA

**SLM:** Speech Learning Model
SP: Sensitive Period

TL: TL

TPR: Total Physical Response

UG: Universal Grammar
Chapter 1 Introduction

1.1 The study

This is a thesis that deals with a study of adult learners of English learning Mandarin tones in an Australian University with a mixture of qualitative and quantitative methodology. The aim of the study is to discover whether certain pedagogic practices are beneficial to second language (L2) learners of Mandarin.

1.2 Problem and Rationale

This thesis revolves around issues of the teaching and learning of Mandarin pronunciation. The emphasis of the teaching method under investigation is upon enabling the acquisition of suprasegmental features such as the intonation, stress, and pauses of whole utterances rather than single syllables. In this way segmental features such as consonants and vowels are expected to develop in a manner more consistent with NS pronunciation.

The teaching method under examination in this thesis combines a number of features, and is described in detail in Chapter 4, Section 4.3. The main aim of this thesis is to explain this teaching method, submit that method to experimental inquiry, make some suggestions as to how the results of the experiment might be interpreted and so make some suggestions for improving teaching of Mandarin to L2 students. The working description of Mandarin phonetics and the literature review that precede the description of this experiment are intended to help the reader come to an understanding of the various salient positions within the field, previous research that has been accomplished, and how some of that research has informed the design of the teaching method examined in this thesis.
Since Mandarin is a tonal language, gaining prosodic competence is arguably more important for intelligibility than it might be in a language such as English. Yet, as we shall see, despite the large amount of research that has been undertaken in Mandarin "tones" very little recent progress has been made in understanding Mandarin prosody from a linguistics standpoint and even less in terms of any relevance such studies have had for pedagogy. This thesis hopes to go some way in addressing these problems.

In pronunciation teaching, most studies deal with the acquisition of segmental rather than suprasegmental features of the language such as stress, intonation and rhythm (in other words prosody). However, Byrne and Walsh (1973) argue that mastery of the suprasegmental features of a L2 is much more important for intelligibility than producing native-like vowels and consonants. Nevertheless, they acknowledge that the former are much more difficult to learn and teach. The teaching methodology described in this thesis addresses these concerns because it seeks to enhance learners’ acquisition of segmental and suprasegmental features of Mandarin prosody simultaneously, enabling students to produce intelligible, NS-like performances of whole words, sentences and phrases, in a very short space of time. In this way, the teaching method under experiment in this thesis contributes to recent directions in FL teaching associated with the spread of "communicative" methodologies.

With the increasing popularity of the communicative approach in language teaching, after all, the focus of attention has moved away from linguistic accuracy and toward conversational fluency. Even the goal of acquiring the sounds of an L2 has changed from the quest to achieve near-native proficiency to more of an ability to develop communicative effectiveness and intelligibility. Thus, according to Nunan (Nunan, 1991), “The major theoretical shift which has
occurred with the development of communicative approaches to language teaching has been from segmental work to a focus on suprasegmental features of rhythm, stress and intonation.” Such observations are relevant for tone languages because phonetically, in a "tonal language," the speaker must work to produce segmental units and suprasegmental features similar to native performance or be rendered communicatively incompetent. In a tonal language such as Mandarin, incorrect use of tone easily leads to confusion or misunderstanding and is a major impediment to effective communication.

In this study “Mandarin refers to “Modern Standard Chinese” is based on the Beijing dialect. Mandarin is China’s official language, known in the People’s Republic of China as Putonghua (common speech) and in Taiwan as Guoyü (national language). It is also the most common Chinese language taught to foreign students of Chinese.

Mandarin has a comparatively simple and limited segmental structure. In Mandarin, sentences are generally simple, short and generally unembellished. Understanding the sentence in the case of errors in prosody relies, more than it does in a language like English, in concrete features of context. In general, this makes prosodic competence even more critical since there is generally a relative dearth of syntagmatic "context" which might otherwise reduce the confusions (and blank faces) that generally attend the incompetent production of tones by non-native speakers (NNS) of Mandarin. Most researchers agree, however, that while the acquisition of segmental features does not present major problems for English speakers learning Mandarin, the acquisition of the tones is consistently the major stumbling block (Guo, 1991). The single major cause of perceived foreign accent in English speakers speaking Mandarin is, in fact, probably the incorrect production of tones (Zhang & Xu, 1981).
While research on the acoustic nature of the Mandarin tones has been extensive, few studies have been concerned with addressing the problems associated with the acquisition of Mandarin tones and prosody by native English speakers. Research on the implementation of the pedagogical measures suggested by these studies, furthermore, has been practically nonexistent. The present study is designed to help fill this gap. Its purpose is to examine the process of acquisition of Mandarin tones and prosody by a group of adult English speakers using a Somatically Enhanced Approach (SEA) to the teaching of Mandarin tones and prosody. The first aim is to determine specifically the nature of their problems in tones and prosody. The second aim is to examine the efficacy of this approach in teaching Mandarin tones and prosody to Australian learners of Mandarin. The practical rationale of this thesis is that the findings of such a study should make possible the designing of teaching approaches - and an effective learning environment - capable of significantly ameliorating these problems.

1.3 Objectives

The present study investigates the process of tone and prosody acquisition of Mandarin of a group of 22 adult native English speakers (undergraduate students taking the first year Chinese program at the University of Canberra in the early stages (first 6 weeks) of their learning.

A major aim of the study is to obtain information about what is happening at the beginning of the process of acquiring Mandarin pronunciation especially with regard to the acquisition of Mandarin tones and prosody. The types of information sought include:

1. Sequence of mastery of the four tones, and hence the order of difficulty of those tones;
2. Common tone and prosodic error patterns among learners;
3. Likely cause(s) leading to the detected errors;
4. How and where interference from English intonation takes place and how significant is its role;

5. Effectiveness of SEA on the teaching of Mandarin tones and prosody;

6. Effectiveness of specific computer enhanced learning programs in the learning process.

1.4 Chapter Outline

This report is organized into eight chapters including the present introductory chapter (Chapter 1).

Chapter 2 presents a brief description of the Mandarin phonological system concentrating on the prosodic aspects of the phonological system.

Chapter 3 presents a review of existing literature on the acquisition of Mandarin tones by English-speaking adults and by Mandarin-speaking children. The review focuses on the specific difficulties encountered and the common errors made by these two groups of learners. Differences in the learning processes of the two groups are discussed in terms of the various theories on the language acquisition by adults and children.

Chapter 4 describes the theoretical framework underlining the Somatically Enhanced Approach to the teaching of Mandarin tones. This includes a detailed description of the classroom methodology adopted in the learning environment.

Chapter 5 describes the design of the experiments involving the experimental and control groups of students based on the data collected from 2001 to 2004. It covers (1) quantitative data collection methods based on students’ oral conversation performances (2) and qualitative data collection methods using end of semester questionnaires and face-to-face interviews.
Chapter 6 reports on the qualitative results obtained in the study.

Chapter 7 describes the quantitative results from both groups of students based on the end of semester questionnaires and face-to-face interviews.

Chapter 8 contains a summary of the conclusions reached in the study and a discussion of the pedagogical implications of the conclusions. Future research directions in this area are also outlined.
Chapter 2 Mandarin Phonological System

In this thesis, Modern Standard Chinese will be referred to as Mandarin from hereon. Prosody refers collectively to variations in pitch, tempo, and rhythm (Crystal, 1997:283). Intonation represents the patterns of pitch or melody (Crystal, 1997:182) and is referred to as the suprasegmental use of pitch on the sentence or phrase level carrying syntactic, semantic, and/or affective meaning. Rhythm may be considered as the perceived regularity of prominent units in speech (Crystal, 1997:302). The rhythmic structure of speech reflects a hierarchical organization of the temporal sequence of speech sounds into syllables and higher-level units of prosodic and syntactic structure. Part of learning a spoken language is the acquisition of its systematic rhythmic organization. Since intonation, tempo, and rhythm are inextricably linked, this study is one of acquisition of tones and prosody, and not of intonation per se.

“Tone” in this thesis is used exclusively to indicate contrastive pitch on the syllable or word level carrying lexical meaning. Mandarin is a tonal language in which pitch is used primarily on the syllable level to signal lexical meaning for every syllable. English, like many European languages, mainly uses pitch to indicate emphasis or stress within a particular word to express syntactic, semantic or affective meaning in a sentence or a phrase. Pitch change in English is not tied to word meaning as in Mandarin. Sentential intonation is also used in Mandarin but its effect on affective meaning is not as significant as in English (please refer to section 2.2.3).

2.1 The phonology of Mandarin

Following the traditional approach to describe the phonology of Chinese, the structure of a Mandarin syllable is presented in terms of initials, finals and tones (Chao, 1968; Cheng, 1973). The following description is based on Li and Thompson’s Mandarin Chinese: A Functional Grammar (Li & Thompson, 1989).

2.1.1 Initials

The initial represents the consonantal beginning of a syllable. Since Mandarin does not have consonant clusters (sequence of consonants), the consonantal beginning of a syllable can only be a single consonant. There are, however, Mandarin syllables that do not have any initial consonants. For those syllables the tradition is to describe their initials as “zero”. The initials of
Mandarin are provided in Table 2.1. Their corresponding International Phonetic Alphabet (IPA) and Pinyin, the national phonetic alphabet adopted in People’s Republic of China is provided in Table 2.2. Including zero initials, Mandarin has twenty-two initials including zero initials.

Table 2.1. Mandarin Initials (consonants)

<table>
<thead>
<tr>
<th></th>
<th>Bilabial</th>
<th>Labiodental</th>
<th>Dental</th>
<th>Alveolar</th>
<th>Alveopalatal</th>
<th>Palatal</th>
<th>Velar</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stop</td>
<td>Unvoiced</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Non-aspirated</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>[p] b-</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Aspirated</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>[pʰ] ph-</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>[t] d-</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>[tʰ] th-</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nasal</td>
<td>Voiced</td>
<td>—</td>
<td></td>
<td>[n] n-</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fricative</td>
<td>Unvoiced</td>
<td>—</td>
<td>[f] f-</td>
<td>[s] s-</td>
<td>[z] zh-</td>
<td>[x] x-</td>
<td>[h] h-</td>
</tr>
<tr>
<td></td>
<td>Voiced</td>
<td>—</td>
<td></td>
<td>[θ] θ-</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lateral</td>
<td>Voiced</td>
<td>—</td>
<td></td>
<td>[l] l-</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Affricate</td>
<td>Unvoiced</td>
<td>Non-aspirated</td>
<td></td>
<td>[ts] ts-</td>
<td>[dz] dz-</td>
<td>[χ] χ-</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Aspirated</td>
<td>[tsʰ] tsʰ-</td>
<td></td>
<td>[dzʰ] dzʰ-</td>
<td>[χʰ] χʰ-</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

(Note: Taken from “Mandarin Chinese: A Functional Reference Grammar” by Charles N. Li and Sandra A. Thompson, 1989, page: 5)
### Table 2.2. Correspondence between IPA and Pinyin Symbols

<table>
<thead>
<tr>
<th>IPA Vowel Symbols</th>
<th>Pinyin Symbols</th>
<th>Context</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>[a]</td>
<td>a</td>
<td>all</td>
<td>lā = [lɐ] ‘pull’</td>
</tr>
<tr>
<td>[e]</td>
<td>e</td>
<td>all</td>
<td>bān = [pɐn] ‘move’</td>
</tr>
<tr>
<td>[i]</td>
<td>ɪ</td>
<td>between [ɪ] and [ŋ]</td>
<td>liān = [liɛn] ‘connect’</td>
</tr>
<tr>
<td>[ʊ]</td>
<td>ʊ</td>
<td>all</td>
<td>yuān = [yɜɛn] ‘fan’</td>
</tr>
<tr>
<td>[ʊ]</td>
<td>ʊ</td>
<td>all</td>
<td>mō = [mʊ] ‘grind’</td>
</tr>
<tr>
<td>[ɤ]</td>
<td>ɤ</td>
<td>before [ŋ] or after [u]</td>
<td>kêng = [kʊɛŋ] ‘dragon’</td>
</tr>
<tr>
<td>[ɤ]</td>
<td>ɤ</td>
<td>all</td>
<td>lào = [lɑʊ] ‘toast’</td>
</tr>
<tr>
<td>[ɤ]</td>
<td>ɤ</td>
<td>all</td>
<td>lè = [lɛ] ‘happy’</td>
</tr>
<tr>
<td>[e]</td>
<td>e</td>
<td>before [t]</td>
<td>tēi = [tɛi] ‘thunder’</td>
</tr>
<tr>
<td>[ɛ]</td>
<td>ɛ</td>
<td>after [t] or [k]</td>
<td>tie = [tɛ] ‘arrange in order’</td>
</tr>
<tr>
<td>[a]</td>
<td>a</td>
<td>after [t]</td>
<td>xiā = [xiɛ] ‘life’</td>
</tr>
<tr>
<td>[ɛ]</td>
<td>ɛ</td>
<td>before [k] or [ŋ]</td>
<td>gēn = [kan] ‘root’</td>
</tr>
<tr>
<td>[ŋ]</td>
<td>ñ</td>
<td>all</td>
<td>gēng = [kʊŋ] ‘tilt’</td>
</tr>
<tr>
<td>[ŋ]</td>
<td>ñ</td>
<td>all</td>
<td>ěr = [ŋ] ‘son’</td>
</tr>
<tr>
<td>[ɛ]</td>
<td>ɛ</td>
<td>after [Cul]</td>
<td>dui = [tʊɨ] ‘correct’</td>
</tr>
<tr>
<td>[a]</td>
<td>a</td>
<td>after [Cul]</td>
<td>dōn = [tʊən] ‘squat’</td>
</tr>
<tr>
<td>[ɨ]</td>
<td>i</td>
<td>after [tɛ], [tɛ̃], [t], [n]</td>
<td>shi = [tʃi] ‘poetry’</td>
</tr>
<tr>
<td>[ɪ]</td>
<td>i</td>
<td>after [t], [tɛ̃], [n]</td>
<td>sī = [ʃi] ‘think’</td>
</tr>
<tr>
<td>[u]</td>
<td>u</td>
<td>after [k], [kɛ̃], [k], [ŋ]</td>
<td>wēi = [kəi] ‘female’</td>
</tr>
<tr>
<td>[u]</td>
<td>u</td>
<td>with any initial except zero</td>
<td>li = [li] ‘drink’</td>
</tr>
<tr>
<td>[y]</td>
<td>y</td>
<td>with any initial except zero</td>
<td>xiū = [yʃi] ‘false’</td>
</tr>
<tr>
<td>[y]</td>
<td>y</td>
<td>after [t], [k], [ŋ]</td>
<td>guì = [kɛi] ‘go’</td>
</tr>
<tr>
<td>[y]</td>
<td>y</td>
<td>after [ŋ] and [t]</td>
<td>tǐ = [ti] ‘donkey’</td>
</tr>
<tr>
<td>[y]</td>
<td>y</td>
<td>after zero initial</td>
<td>nǚ = [ŋi] ‘daughter’</td>
</tr>
<tr>
<td>[y]</td>
<td>y</td>
<td>after zero initial but not in isolation</td>
<td>yī = [tj] ‘fish’</td>
</tr>
<tr>
<td>[y]</td>
<td>y</td>
<td>after zero initial but not in isolation</td>
<td>yuān = [yɛn] ‘garden’</td>
</tr>
<tr>
<td>[ʊ]</td>
<td>ʊ</td>
<td>after zero initial but not in isolation</td>
<td>yào = [lʊəu] ‘want’</td>
</tr>
<tr>
<td>[ŋ]</td>
<td>ŋ</td>
<td>in isolation</td>
<td>yī = [ŋi] ‘one’</td>
</tr>
<tr>
<td>[u]</td>
<td>u</td>
<td>in isolation</td>
<td>wēn = [ʊɛn] ‘smell’</td>
</tr>
<tr>
<td>[u]</td>
<td>u</td>
<td>in isolation</td>
<td>wǔ = [u] ‘five’</td>
</tr>
</tbody>
</table>

(Nota: Taken from “Mandarin Chinese: A Functional Reference Grammar” by Charles N. Li and Sandra A. Thompson, 1989, page: 7)
2.1.2 Finals

The final is the part of the syllable excluding the initial. There are thirty-seven finals in Mandarin, as shown in Table 2.3 in IPA symbols.

Table 2.3. Mandarin Finals (vowels)

<table>
<thead>
<tr>
<th>Finals</th>
</tr>
</thead>
<tbody>
<tr>
<td>1, 2, 3</td>
</tr>
<tr>
<td>i</td>
</tr>
<tr>
<td>u</td>
</tr>
<tr>
<td>y</td>
</tr>
</tbody>
</table>

(Note: Taken from “Mandarin Chinese: A Functional Reference Grammar” by Charles N. Li and Sandra A. Thompson, 1989, page: 6)

The velar nasal, [η], occurs only as part of a final, never as an initial. In Pinyin, it is represented by ng. The finals, as can be seen from Table 2.3, are composed mainly of vowels. The only two consonants that occur in a Mandarin syllable final are the velar nasal, [η], and the alveolar nasal [n], and these may occur only at the end of a final.

In the early 20th century, the tone and intonation research for Mandarin Chinese entered a new phase mainly due to the work of Dr. Chao Yuan-ren. Chao’s pioneering work provided Chinese linguistics and the field of teaching Chinese as a FL a description of Mandarin phonetics which has ever since been followed by both phoneticians and practitioners of Mandarin language teaching and research (Chao, 1930). As this particular description of Mandarin phonetics has become the norm in Chinese linguistics and the field of teaching Chinese as a FL (FL), it follows that in this thesis, Chao’s description is used to discuss the relevant issues associated with tone and intonation research.

Chao’s description, in fact, is not the only system of description available. There are alternatives, but none are in such wide use. It would also be possible to create a whole new
system of description, as many researchers do. No such system has been created for this thesis, however, for two main reasons. Firstly the benefits of any improvements to description that might be reaped are likely to be outweighed by the confusion created by forcing readers to learn a whole new system of representation. Secondly, this thesis is not about quibbling over accuracy of a description of Mandarin phonetics which has so far proven relatively serviceable.

The following description of the phonetics of Mandarin has been developed from the work of Li & Thompson (1989) and Guo (1991).

2.1.3 Tones
According to Chao's system, Mandarin has five tones; a level, a rising, a falling-rising, and a falling tone, plus a neutral tone. All of these tones are loosely tied to five relative pitch levels (see Table 2.4). As explained in this thesis, in order to depict tones graphically for ease of understanding, the “Notational System of tones” (Chao, 1930) is used to represent Mandarin in terms of five tones. This thesis is not a defence of Chao's system. Instead it takes Chao's system as a construct with which it is possible to work. In any case, in this thesis, Chao's notation is not a set of "lenses" through which samples are perceived but merely a system enabling descriptions of samples to be easily understood by readers educated in Mandarin. It will be noted that given the complexity of Mandarin tones, let alone prosody, no verbal description by the teacher can be expected, by itself, to significantly aid students in the production of tones. This will become increasingly obvious as we examine the research.

2.1.4 The Notational system of tones
The Notational system of tones use in this thesis is a well-known system called the tone letter system (Chao, 1930). In the early part of the twentieth century, Chao (1930) proposed the use of five different letters to express the tonal value of different lexical tones. 1, 2, 3, 4, and 5 will be used for this purpose:

“1” expresses the lowest tonal value,
“2” expresses the next lowest tonal value;
“3” expresses the middle tonal value,
“4” expresses the next highest tonal value; and
“5” expresses the highest tonal value.

In most textbooks, the tones are graphically represented by the *Figure of Chinese Tones* as shown in Table 2.4. Emphatically, each number does not denote a particular absolute frequency value and the difference between the five different levels does not correspond to any standardized absolute frequency value. They only reflect the contrasts between highs and lows. This is what is usually known as the “tone letter” notational system for tones. As aforementioned, the tone letter system of tone notation has become the most commonly used notational system with respect to teaching and research in the study of Mandarin phonology. Though not perfect, this system of discussing tones is most commonly used and understood by researchers in the field. Therefore, discussion of issues with reference to this tone letter system will make the discussion most accessible to participants. In order to communicate to wider audience, it has been decided that this system is adopted for this thesis. When teaching Mandarin using this notational system, it is referred to as the *tone shape method*.

Table 2.4. Figure of Chinese Tones

An important feature of this notational system is that it mainly relies upon the perceptive experience of NSs to judge the difference between the different tones. The determination of
each tonal level depends on the height difference between the beginning of a tone and the end of a tone. For instance, in Mandarin,

Tone 1: The first tone is usually read as high compared to other tones (e.g. 妈: mā: mother, the value for this tone can be then assigned as 55;)

Tone 2: The starting point of the second tone is neither high nor low and its end point ends on nearly the same height as the first tone, so its tonal value can be assigned as 35 (e.g. 麻: má, hemp). One of the most interesting phenomena involving tones in the Chinese dialects is called tone sandhi. Tone sandhi may be described as the change of tones when syllables are juxtaposed. To put it differently, a syllable has one of the tones in the language when it stands alone, but the same syllable may take on a different tone without a change in meaning when it is followed by another syllable. A tone sandhi rule in Mandarin involves the second tone, which changes into the first tone when it is preceded by either the first or the second tone and followed by any one of the four tones.

Tone 3: The third tone falls first and then rises. Its starting point starts lower than Tone 2, therefore this can be assigned as a value of 2. It then falls to the same lowest level as the fourth tone, we can assign this value to 1, then it rises to the medium high level, and this can be assigned as a level 4. So the tonal value of this tone is 214 (e.g. 马: mă: horse)

Two more important tone sandhi rules in Mandarin involve the third tone:

Tone sandhi rule 1: When a third tone syllable is followed by a syllable with any tone other than another third tone, the third-tone syllable changes to a low-tone syllable with the pitch contour 21. For example, 妈 “horse” has the third tone in isolation, but when it is followed by another syllable such as 车 “vehicle”, the sequence is pronounced with following tone sequence: 妈(21) 车(55).

Tone sandhi rule 2: When a third tone syllable is followed by another third tone syllable, the first one changes into a second tone. For example, 赶 “to chase” and 鬼 “demon” both have
third tones. When they are in sequence, gān guī “to exorcise demons”, gān is changed from third tone (214) to the second tone (35).

Tone 4: The starting point of the fourth tone has the same height as the first tone, so a 5 can be assigned to its starting point. However, its end point is the lowest, therefore can be assigned as 1. So for the fourth tone, from the highest to the lowest, the tonal value can be assigned as 51 (e.g. mà, to scold)

Neutral tone: If a syllable has a weak stress or is unstressed, it loses its contrastive, relative pitch and therefore does not have one of the four tones described above. In such a case, the syllable is said to have a neutral tone. According to Chao (1968), the pitch of the neutral tone is:

- Half-low after first tone: tā-de-his;
- Middle after second tone: hóng-de- red one;
- Half-high after third tone: wō-de- my;
- Low after fourth tone lǜ-de=green one.

(Neutral tone in this thesis is indicated with a number 5 after the syllable. In some books, the neutral tone is represented by 0 rather than 5)

The segmental part of the Chinese words is represented by the widely known and accepted romanization system called Pinyin. Pinyin romanization system is the standard adopted in this thesis.

2.2 Mandarin prosodic system

There have been many studies on the tonal system of Mandarin Chinese (Dreher & Lee, 1966; Hombert, 1975; Howie, 1976; Kiriloff, 1969; Tseng, 1981). However, almost all the research or description on the tonal systems of Mandarin Chinese concentrated on either lexical tones or pairs of tones. In the next sections, two descriptions of the Mandarin prosodic system will be discussed. Only one description looks at tones in running speech (Kratochvil, 1998). Another came from research done on prosody modelling for speech recognition purposes (Wang, 2001).
2.2.1 Fundamental frequency, duration, amplitude and tone

Kratochvil’s description on the prosodic system of Mandarin is based on the Intonation in Beijing Chinese (Kratochvil, 1998). The description is based on the discussion of Beijing Dialect (BD) which is the basis for Modern Standard Chinese (Mandarin). Based on data collected on running speech rather than isolated syllables, each syllable of the corpus used in Kratochvil’s data was treated as a carrier of the segment of the laryngeal voice output coinciding with its tone domain as defined in Howie (1974). Within the complex of features reflected in the prosodic shape of the BD breath group (the prosodic complex), the smallest distinctive configuration coincides with the syllable. This configuration traditionally is referred to as the tone. Both Kratochvil (1998) and Wang (2001) concur that tone is primary in the sense that larger configurations up to and including the prosodic complex itself are specific modifications of a string of one or more tones.

A tone is polydimensional. As the product of a complex laryngeal gesture, it can be described fully in acoustic terms only by referring to the relationship between all the dimensions of the laryngeal voice, that is, the respective variation of fundamental frequency (F₀) and amplitude (A) over varying duration (T). It has been demonstrated experimentally, in Rose (1984), that A variations in Chinese tones cannot be explained as functions of F₀ variations, and that A variations are used as important cues in tone identification. Because of this and the primary function allocated to the frequency in the acoustic dimension of prosodic features, still widely held in phonetic research, it is no longer tenable for BD.

It should be apparent from the above discussion that Chao's description of Mandarin tones described in Section 2.1.3 of this chapter cannot possibly be taken as a complete representation of any tone, but only as an abstract notation. Given that Chao's notation cannot be taken as an "accurate" representation of Mandarin, there cannot be anything very wrong with L2 beginners in Mandarin creating their own notation which for the student, at least, actually "looks" like the sounds they are perceiving. However Chao’s notation of the four tones (later translated into the four tone diacritics) at least indicates the contour of the four tones more accurately than perhaps a student’s own invention. So while students’ created romanization might look like the sounds they perceive, diacritics could be introduced to students to help them visualize the tones. Later
in the course of learning Mandarin, they will need to learn romanization and Chao's system of notation in the form of tone diacritics in order to participate more fully with respect to the cultural and academic opportunities created via their study of Mandarin.

2.2.2 Word and phrase features: stress

In suprasegmental units, BD tones are modified in strings of two or more syllables according to the requirements of prominence (Yang, 1994). Prominence means the stress or emphasis by which a particular part of a speech performance is drawn to the attention of the listener. Prominent syllables are not only intrinsic to the rhythmic patterns of a given language (such patterns might be said to belong to the language itself); they are often used by speakers, intentionally or unconsciously, to facilitate communicative effects. Prominent syllables are generally louder; their pitch level is higher (in the case of tones characterized by a $F_0$ peak). They are also of longer duration, especially in the case of Tone 3. The basic stress patterns in BD, from which all others can be derived by observing the rhythms of speech, are unstressed-stressed (US) and stress-unstressed (SU). In the unstressed to stressed pattern, the first syllable tends not to be affected by prominence either positively or negatively, while the second tends to be prominent. In the stressed to unstressed pattern the first syllable tends to be prominent, and the second not affected by the desire of speakers to give prominence to one syllable or another.

In traditional, non-experimental descriptions of the prosodic features of BD syllabic strings, categories of prominence and stress are simply not recognised. Explanations of all secondary modifications are described in terms of the presence and absence of tones and of their interplay. This is because of the force of rules based on impressions of isolated citation forms of tonal patterns that have become part of the descriptive and pedagogical convention in Chinese linguistics. In normal BD speech, most of these rules do not apply at all, or at least not in the categorical way in which they are formulated. As Kratochvil (1998) noted whatever modifications in the prosodic features of tone contact are actually made in normal speech cannot be explained without reference to prominence patterning. Tone variation cannot be separated from secondary prosodic modifications, and also from segmental phenomena.

Again, from the previous discussion, it will be seen that it is very unlikely that attending to tones in isolation can be expected to dramatically improve student's intelligibility in actual
conversation, even if they could somehow manage to consciously remember what the tone of a particular lexical item is (which a few L2 speakers of Mandarin, admittedly, manage to do -- going a great way to maintaining the prestige associated with mastery of such a supposedly "difficult" language). While teacherly corrections, explanations and modelling can certainly help in the Somatically Enhanced Approach described in this thesis, these are not the most important ways in which students are assisted in their attempts to become competent producers and interpreters of Mandarin -- a task that simply cannot be achieved through disregard for tones and other aspects of prosody.

2.2.3 Tone and Intonation

Furthermore, despite the great deal of work done on the relation between Mandarin tones and other forms of intonation, there has been no consensus, even among the experts as to the effect of intonation on tone. A question, by and large, has been whether tones are pristine models or forms over which other intonation patterns are grafted or whether they are of the same epistemic order as other patterns of intonation. The ramifications of such a question for pedagogy remain unclear and are likely to remain so.

In the literature on the interaction between tone and intonation of Mandarin, two opposite assumptions exist. The pitch-neutralization position (Ho, 1976; Rumjancev, 1978; Tseng, 1981; Wu, 1982) claims that the $F_0$ values of the four citation tones are affected by the sentence intonation within two "domains", "shape" and "scale". This is to say that a tone is a sound with a specific kind of contour (which can be made at a variety of volume levels). These studies all conclude that despite any obvious changes to contour resulting from other types of intonation, the basic contour features of these four tones do not change. Although intonation causes some modifications of tone contours, they do not, according to the proponents of the pitch-neutralization position become "unrecognizable". Because the utterance remains intelligible, the argument goes the tone then must have been recognized. Therefore it could not have possibly changed.

This circular argument involves a fallacy which, in turn, derives from mystification. The tones as Chao described them are now being imagined as things in themselves. It remains to ask, of course, whether the tones become unrecognizable as a result, or whether it is simply the case
that the sentence remains intelligible as a result of other features (or other contextual features) being recognised. Even so, the pitch-neutralization position retains some merit. For NSs of Mandarin, tones do seem to have some sort of independent force which impacts upon the shape of syllables regardless of the mood or register of the utterances. With regard to tone scale, however, all investigators affirm that the tonal values of the four citation tones are indeed affected by sentence intonation. Question-intonation, for example, raises the tonal value whereas statement-intonation lowers the tonal value (especially those utterance-final syllables).

Chao (1933), however, had a different opinion. According to his pitch-prolongation position, both the tone shape and tonal value - the contour, are so affected that the tones become unrecognizable. Chao also does not think the sentence has become unintelligible. It is just that the contour has significantly changed. Note that the advocates of both positions are not disagreeing with each other. In a way, the dispute between these two positions is thus largely unnecessary. All agree that the shape and scale of tones are affected by other forms of intonation and all agree that this modulation does not have much an effect on communication.

Shen (1989a) also conducted an experiment to investigate further the interaction of tones and intonation in Mandarin. Her experiment demonstrated that intonation does have an effect on lexical tones, changing tonal values. Interrogative intonation, Shen found, raises not only the tonal value of the sentence-final syllable but also the overall tonal value of the whole utterance. Shen (1989a) thus asserts that intonation-tone interaction in Mandarin involves a pitch change in both the shape and scale of a given tonal pattern. Nevertheless, the basic features of all tones, according to Shen, remain intact. For example, the falling tone does not become falling-rising tone under question intonation. When Mandarin sentences are perceived by NSs in her study, the effect of intonation on tones is not so great to render the tones unintelligible. Shen (1989a) also indicated that tones at sentence initial and final positions seem to behave differently from at other positions.

Shen is clearly trying to find some synthesis between the pitch-neutralization and pitch-prolongation positions. Her study offers independent findings which are not always in agreement with those of other researchers (Chao, 1933; Ho, 1977; Rumjancev, 1978; Tseng,
1981; Wu, 1982). Of course, language variations being what they are, any or all of these studies might be accurate.

Regardless, there is little benefit to be had by coming to a position regarding whether or not tones always remain "recognizable" in NS performances of Mandarin. The dispute that lies between the pitch-neutralization and pitch-prolongation positions may be traced from the distinction Chinese linguists make between tone and intonation. Following Chao (1930), tones are first used to describe language but soon become quasi-platonic forms which seem to take on a real existence of their own. No longer merely heuristic constructs, it then becomes possible to abstract tones from the intonation patterns of real speech and to dispute whether or not they survive in their pure form regardless of all other intonation no matter how significantly such factors might objectively alter F0.

If it were not the case that tones do seem to have some sort of independent agency, the dispute might be easily dismissed. As it stands it seems wisest to keep an open mind on the subject and attempt to proceed, pedagogically, without presuming either the real existence of tones independent of intonation per se or otherwise. For teachers, the relevance of extant research into the relation between tones and intonation might well lie more in exposing some of the reasons for the comparative lack of progress in Mandarin teaching than in pointing out viable pedagogical directions based upon those findings.

A few studies have been conducted on the interaction between tone and sentence focus (Gårding, 1987; Xu, 1999). These studies found that the lexical tone of a syllable is the most important factor for determining the local fundamental frequency (F0) contour while the demands of prominence extensively modulate the global shape of the F0 curve. In other words, they are in substantial agreement with everybody else, but add in a bit of worthwhile, precise detail as to how this actually happens. The effects of focus are asymmetric: the F0 range of words at a non-final syllable which is being given prominence is substantially enlarged, especially the high value. The F0 range after such a syllable is both lowered and reduced while the F0 range before the focus remains similar to cases where there is no such focus.
Downstep and declination are important aspects of intonation. Downstep refers to the phenomenon that a high (H) pitch target has lower $F_0$ height after a low (L) pitch target; while declination refers to the tendency for $F_0$ to gradually decline over the course of an utterance. A broad term “downtrend” is used in this thesis to describe the combined effects of the two. It is argued (Liberman & Pierrehumbert, 1984; Pierrehumbert, 1980) that the declination in English is the outcome of the downstep of subsequent H pitch accents throughout the utterance. However, a study of Mandarin Tone 1 syllable sequences (Shih, 1997) has found that Mandarin has a strong declination effect regardless of intonation. The $F_0$ decline can be treated as an exponential decay, and the slopes of the decline are influenced by sentence focus as well as by the length of the sentence.

A great deal of work has thus been done on the effect of intonation upon tone, and many worthwhile insights have been made, contributing greatly to the fields of linguistics and phonology. Regardless of whether one agrees with the different theories of tone and intonational interaction, there is a ramification for pedagogy here. A teacher who sides with the pitch-neutralization position is more likely to imagine tones as glass sculptures in the sky and to endless drill his/her students in attempts to imitate those pure forms. A teacher who sides with Chao (1933) is likely to see that tones change according to a vast plethora of uses and so more likely to simply throw up his or hands in defeat, lacking any practical method by which to approach the problem. Since even the experts cannot agree on how tones change their contours as a result of other patterns of intonation, simply telling students about such changes cannot reasonably be held forth as a very sound pedagogical approach.

SEA does not rely on analyses, categorizations or descriptions of such modulations and does not proceed by explaining these shifts in contour to students in the assumption that they will consciously monitor their own production accordingly. Instead, SEA provides students with authentic materials, replete with such variations, and enables students to interact with those materials in such a way as to effectively sidestep the sort of disputes outlined above; and to duplicate the intonation of such materials through practices that are not wholly reliant upon conscious monitoring.
2.3 Summary

This thesis attempts to develop a teaching pedagogy for teaching the prosody of Mandarin which includes tones and their primary and secondary modifications. Its intention is not to investigate the validity of any Mandarin intonation theory mentioned above. Furthermore, the experiments conducted in this thesis were not specifically designed to resolve the conflicting views on the interaction between tone and intonation. However, one of the research areas is to investigate how the SEA affects the perception and production of Mandarin prosody. Therefore, the investigation concentrates upon how L2 students’ prosodic systems in Mandarin develop. In the next chapter, a literature review concerning first and SLA; first and SLA of Mandarin and other related research will be conducted.
Chapter 3 Review of Literature

3.1 Tone Acquisition by Mandarin-speaking children as a L1

3.1.1 Studies on the acquisition of Mandarin tones

In the few studies on the acquisition of Mandarin tones that have so far been undertaken, researchers concentrated their effort on accounting for the sequence of tones produced by various groups of first language (L1) children. In other words, these studies are concerned with whether little Chinese children master tone 1 before tone 2 and so on. As a result of such studies certain curricula have been developed, informed by one study or another. Furthermore, these studies are not in substantial details as to the actual order in which L1 children learn tones, agreeing only that there must be some such "natural" order.

It is questionable, moreover, whether the developmental sequence of tone acquisition significantly reflects the general features of cognitive and perceptual processing that are exhibited by L1 Mandarin speaking children. In other words, L1 children might or might not have mastered all the tones for listening purposes even before they have uttered their first tone to the satisfaction of their auditors. The aim of such research, after all, was not to investigate what L1 Mandarin children do when these children encounter a stream of spoken Mandarin input. It was merely to say which tones are learned in which order. Consequently, how L1 Mandarin children deal with syllables, words and sentences is still not clear. Nor it is apparent what role tones actually play in the act of segmenting a stream of spoken Mandarin input for Mandarin children. Even less clear is to what relevance the order in which tones are acquired by L1 children might have for adult L2 speakers seeking to acquire Mandarin prosody and tones.
There is, furthermore, little cross-over between the body of research discussed in the previous chapter of this thesis and the research here under discussion. As such, these researchers have assumed that tones are self-evident features of language which are presumably tied to some generative structure. The category of tones indicated by the tone labels introduced by Chao (1930), has been adopted as the norm in Chinese teaching and linguistic circles. These studies do not diverge from that system. While not a problem in most situations, for this sort of study it is a fatal flaw. Making the assumption that there are four (or five) tones and then finding out in what order they are acquired by L1 children is not only beginning one's study in a state of theory-laden observation. It inscribes the structure of a supposedly pre-known, and self-evident, object into the form of the inquiry even before approaching the actual phenomena. This is a bit like looking for animals and having a form with headings - "lion," "tiger," "bear" - and so treating anything else coming out of the forest as ipso facto not an animal. These studies claim to observe children but any production that does not fit their pre-arranged schema is not counted. They should have examined their samples first and then produced a system of notation appropriate to the actual data. But even when we accept these studies on their own terms, and try to learn from them rather than dismiss them, it remains unclear whether tone categories play any role in the children’s attempts to segment a stream of Mandarin input. As far as pedagogy goes, reason for the research seems to be trying to determine which tones should be taught to adult L2 speakers first. This position, however, is in complete disregard for any urgencies students may have in terms of carrying on any real conversations or performing any real life tasks (e.g. buying a train ticket).

The motivation for the undertaking of such studies, historically, was probably because developmental sequences in segmental phonology had been conducted extensively. Some
Chinese researchers conducted research in segmental phonology and then proceed to search for developmental sequences in suprasegmental phonology (Clumeck, 1977). Yet as if living on a desert island ruled by tones, Chinese linguists engaged in these studies seem, ironically, completely isolated from the main trends in similar studies carried out with respect to different languages.

Studies into L1 acquisition on infants show that there is a universal tendency, regardless of TL, for learners to segment words in response to stress patterns. Peters proposed the first principles of segmenting prosodic features in an auditory stream: (1) segmenting stressed syllables; (2) segmenting at rhythmically salient places; and (3) segmenting at intonationally salient places (Peters 1985, cited in Jusczyk, 1998). This might also be true in the acquisition of Mandarin as an L1 for children. However, no research investigating the phenomenon of segmentation has been done in the area of L1 acquisition of Mandarin so as to provide us with any insights in how Mandarin speaking children segment the acoustic stream in Mandarin. In other words, studies carried out in L1 acquisition bear very little resemblance to studies carried out in Mandarin L1 acquisition.

Even so, despite the lack of research in the area, insights gained into the process of tone acquisition by Mandarin speaking children could still shed some light on the difficulties encountered by foreign adult learners in learning Mandarin. As noted, only a limited number of studies on the acquisition of Mandarin tones by children have been carried out. The following review describes work done in this area.

The first systematic study of the acquisition of the Mandarin tone system by children was carried out by Li and Thompson (1977). They investigated the spontaneous speech of 17
Mandarin-speaking children aged between 18 and 34 months over a period about 7 months in Taipei, Taiwan. Each child was visited every two or three weeks. The main method used to collect data was to show pictures in the children’s picture book and to ask each child to name the objects in the pictures. If the child did not respond, the person conducting the test would often supply the answer, which the child might then imitate. They found that the tone system was correctly acquired relatively quickly and was mastered well before the segmental system. They reported that Tones 1 and 4 were acquired earlier and more easily than Tones 2 and 3. The confusion between Tones 2 and 3 was still present when the children reached the 2 to 3-word stage. They also indicated that the tone sandhi rules were learned as soon as the children could produce multiword utterances and the neutral tone was often misinterpreted as a full tone. In Chapter Two of this thesis, section 2.1.4, neutral tone is defined as a weakly stressed or unstressed syllable which has lost its own contrastive pitch. Syllables with neutral tones usually occupy the second syllable or last syllable in a two or three syllable compound. In adult speech, the production of the neutral tone is dictated by the tones of the first syllable in the compounds. For instance, in adult speech bàba, the second syllable ba loses its original contrastive Tone 4 when pronounced in the compound bàba. However, in these children’s speech, a compound such as bàba was produced as bàbà (i.e. ba’s original full Tone 4).

Li and Thompson (1977) reported that the two youngest subjects did not produce Tone 2 or Tone 3. One of them produced no contour tones at all but substituted level tones of different heights for Tone 2, Tone 3 and Tone 4, while producing Tone 1 correctly. The other child produced only Tone 1 and Tone 4 and refused to repeat any Tone 2 and Tone 3 words, which Li and Thompson (1977) suggested that it was because these tones were too difficult for her to produce at that stage. Another explanation might be that this child did not know how Tone 2
and Tone 3 sounded and consequently refused to produce them. The cause of non-production might be a lot more complicated than just the perceived “difficulty” attributed to the tones by the researchers.

Li and Thompson (1977) therefore claimed that Tone 1 and Tone 4 are easier for children to learn to produce than Tone 2 and Tone 3. To support their claim they cited from literature on production and perception studies on adults (Ohala and Ewan Ohala, 1973; Sundberg, 1973). These studies indicate (1) that falling pitch is executed faster than rising pitch, suggesting that greater physiological effort is required in the production of a rising pitch; and (2) that falling tone patterns are perceived more accurately than rising tone patterns, suggesting that rising tones are more difficult perceptually than falling tones. However, this finding is not supported by other studies on tonal languages. Li and Thompson (1977) noted that the acquisition order of Thai tones reported by Tuaycharoen (1977) was that the falling and high level tones are acquired last in Thai but first in Mandarin; the rising tone is acquired earlier than the high level and falling tones in Thai, whereas it is acquired last in Mandarin. None of the studies cited above conducted perception tests to see at this age what can be perceived by the tones. Without such data it is difficult to assign any value to the data gathered because the greatest obstacle to production might be the under development of the children’s vocal apparatus.

Clumeck’s phonological profile on the child M is the only study that offers glimpses of how a Mandarin speaking child might develop their word segmentation and recognition skills in Mandarin (Clumeck, 1977). For the child M, pitch was initially used to signal differences in mood and emotional state at the age of 1; 2 (one year and two months in age) to 1;5.2 (one year, five months and 2 days in age). From 1;5.2 to 1;10, both segments and pitch had been used for
communicative functions. Only one month later, at 1;11, M started to use words and pitches that resembled adult’s use of words. At this stage, a system of tones began to emerge with Tone 1 and 4 produced correctly and words with rising pitch in the adult language were consistently produced by M with a rising pitch.

It was also at 1;11 that M started to imitate words with adult models. However, he had the tendency to produce words containing dissimilar syllables with either one syllable only or by making both syllables alike in sound structure. That is, he would pronounce the adult form “qīngwā” (frog) as “[jā]. From the age of 2;1 to 2;5.2, M made use of this strategy on 41 occasions, and on 36 of these, he preserved the second syllable. Clumeck (1977) attributed this to the fact that the final syllable in Mandarin tends to receive heavier stress. In L1 acquisition study, based on cross linguistic data, Dan Slobin proposed a number of operating principles to explain how a child might segment a stream of sound in a given language. One of the first operating principle which is also known as operating principle A is that children would “pay attention to the ends of words” (Slobin, 1973:191). M’s behaviour of stressing the second syllable in a two syllable compound might constitute limited evidence for Slobin’s operating principle A.

Clumeck (1977) argued that although rising pitch is more difficult for adults and takes longer to execute than falling pitch, it is unknown to what degree it is harder for a child. Clumeck (1977) concluded that it cannot be determined if any tones are more difficult than others for children, and that it is not possible to arrive at a universal hypothesis about the sequence of tone acquisition until more studies have been undertaken. Jeng (1979, 1985) agrees with Clumeck (1977) suggesting that Mandarin tones are acquired by children in no particular order.
Moreover, in the case of boy M in Clumeck’s study (Clumeck, 1977), the parents frequently produced Tone 1 words in Tone 2. Perhaps this caused Tone 2 to be the first tone acquired by M. Based on such results, Clumeck (1977) proposes that the results might simply imply that the frequency of certain prosodic cues in the input is likely to be more instrumental for the acquisition of a tone than any particular innate degree of difficulty measured in terms of the physiological effort required for producing the tones.

Since there are almost no studies in the provision of input in Mandarin to L1 or L2 Mandarin children, it might be pertinent to draw upon the researcher’s limited experience of educating her eldest daughter to illustrate further possible effect of input on acquisition:

The researcher attempted to teach her eldest daughter, Xinxin, Mandarin in the first three years of her life. The child’s father is Australian who speaks less than 20 words of Mandarin. During the first three years of her life, her mother spoke to her in Mandarin only and her father to her in English. When she was a few months old (before she was able to verbalise any sounds), a portion of a typical day’s Mandarin language input from her mother included:

Scenario 1:

Researcher: Xīnxīn yào bù yào hē tāng?

(Xinxin do you want some soup?)

Xinxin smiled and the researcher picked up the bowl of soup and then said:

Researcher: hěn tāng, māmā chuī chuī.

(The soup is very hot. Mum will blow on it.)
(This was followed by a blowing motion by the mother.) Then the researcher fed Xinxin the soup.

Scenario 2: (on the same day)

Researcher: Xīnxīn yào chī tán găo ma?

(Does Xinxin want some sweets?)

Xinxin smiled and accepted.

Scenario 3: (at night)

Researcher: Xīnxīn xiànzài tāng xià shuìjiăo.

(Xinxin lie down and sleep)

(This was usually accompanied by putting her down in her cot and then singing a lullaby to send her to sleep.)

In these three scenarios, the same syllable “tang” appears in four different forms each with a different pitch (i.e. tone). That is “tāng” means soup, “tàng” means “it is hot in temperature”. “Tāng” means to lie down and “táng” means “”sweets”. In each scenario “tang” was presented and then feedback was provided back to Xinxin through either action or tactile reactions. The fact the same syllable “tang” appeared so frequently in her input but with different tones would indicate to her that the changing pitch of the word “tang” was essential for meaning differentiation. It is then not surprising that for a child who is learning to segment words in such
situations to realize that pitch is important in differentiating “tāng(soup)” from “tàng(hot)”, from “tāng(to lie down)” and to “tāng(sweet)”.

The above anecdote illustrates how input could contribute to L1 Mandarin children’s acquisition of Mandarin tones:

Input can be responsible for convincing infants that tones are important in meaning differentiation early in life.

Input could be responsible for certain tones being produced early and more frequently by L1 Mandarin speaking children.

The way input is provided can contribute to the acquisition of certain tones.

Furthermore, in the above scenario with Xinxin, there was a lack of pronouns used to address either the researcher (who is the mother) or the child (Xinxin). In each case the position within the family structure (māma: mother) and the child’s name were used. Since “māma (mother) and “bàba” are likely to be the first words learned by children, Tone 1 and 4 would be likely to make an earlier appearance than other tones. This contrasts sharply with the situation where L2 adult learners are learning Mandarin as a FL. In the FL context, the preservation of learners’ identity, self-concept and motivation is vital. Therefore, “I” language which starts with “wǒ (I)” and pronouns need to be introduced first as the communicative and social needs of adults are so much greater than that of children. Therefore, in a FL context, the appearance of Tone 3 through the frequent use of “wǒ:I” and “Nǐ: you” could be the main reason for the different hierarchy of tones observed.
Another reason for the earlier acquisition of Tones 1 and 4 might have to do with the way input is directed at infants. Papousek et al (1991) conducted a study on melodies in motherese in Mandarin Chinese. From the data of 10 Mandarin mothers, she collected video and audio-recorded data during spontaneous interactions with their two-month-old infants. The data showed that in terms of lexical content, Chinese mothers uttered a significantly high proportion of modelling sounds. Furthermore, in terms of the average acoustic structure of prevalent melodic contours, Chinese mothers allowed the pitch to vary only within a limited frequency range for purposes other than lexical contrasts. In Chinese maternal speech, the neutral tone particles were lengthened and became the carrier of a major part of sentence intonation. At the same time, smoothing or flattening of lexical tones were used as simplified, exaggerated global intonation contour.

In another study investigating how different registers were used to talk to pre-syllabic infants conducted by Papousek and Hwang (1991) shows that in baby talk to presyllabic infants, Mandarin speakers raised peak and minimum $F_0$, reduced $F_0$ fluctuation rate, and increased the proportion of terminal rising contours. Speakers reduced, neglected, or modified lexical tonal information in favour of simplified and clarified intonation contours when talking to Chinese infants. This again reinforces the possibility of why Tone 1 and Tone 4 were acquired first in many L1 students who speak Mandarin when they were young.

Therefore, it is thus entirely plausible that it is input which plays the main role in organizing the development of a “sense of tone” (Guo, 1991) in Mandarin children. As for conclusions on the natural order of tone acquisition for Mandarin, all one can safely state, with these limited results from such a small number of subjects, is that no reliable generalizations can be made. Rather, in
practice, the order of acquisition seems to be determined by what the child is exposed to (Clumeck, 1980; Yue-Hashimoto, 1980). As such, since there does not seem to be any natural order in which L1 speakers of Mandarin learn tones as children, there seems to be no reason to assume that any given tone will be so much "easier" for L2 learners of Mandarin to learn or produce. In other words, the tones can be learned in any order. Therefore, there is no rational reason not to dispense with endless "preliminary" tone drills before they can do any more damage to student motivation. In Chao’s Mandarin Primer: In Intensive course in Spoken Chinese (Chao, 1972), Chao unfortunately saw audio lingualism as the only way to teach learners to perceive Mandarin sounds and thus recommended the first 100 hours to be devoted to such drills.

These studies have no particular relevance to the design of SEA which does not involve teaching tones in any particular order and gives priority, in terms of the selection and order of materials within a curriculum, to student's desire with respect to the usefulness of the utterances.

3.1.2 Relationship between perception and production

Little data is available on the relationship between perception and production in L1 Mandarin children’s language acquisition. As noted, Clumeck ‘s investigation of children’s perception of tones appears to be the only published study of tone-perception in children acquiring Mandarin (Clumeck, 1977). In this study, children P and J served as subjects in a set of tone discrimination tests. The purpose of the tasks was to determine which tonal contrasts the children could perceive. At each test session, each child was given three objects, or pictures of objects, of which two had names that were identical except for the tone, while the third had an entirely different name. The experimenter would pronounce the names of the objects used in the
test and then asked the subjects to identify each of the three objects five times in random order. The results showed the children experienced greater difficulty in learning to discriminate between Tone 2 and Tone 3 than between any other pair of tones. This is in agreement with the data from the tone production tests conducted in his other experiment. This seemed to suggest that it is necessary to perceive before you can produce. An alternative explanation could be that young infant or children learning Mandarin as a L1 especially in the first two years have not sufficiently developed physically or to produce the contour tones Tone 2 and Tone 3.

Similarly in an experiment regarding perception of tone for adult NSs of Mandarin conducted by Chuang (cited in Guo, 1991), subjects mistook Tone 3 as Tone 2 twice as much as they mistook Tone 2 as Tone 3. This might be because of the complexity of tone-sandhi rules with regards to Tone 3. L2 learners equally had problems perceiving Tone 2 and Tone 3 as Chinese nationals do due to Tone 3’s complex tone sandhi rules. Research findings from Kiriloff (1969) also confirmed this point. In Kiriloff’s experiment, 12 subjects who speak Mandarin as a L2 were used in a differentiation test through hearing. He found that errors with Tone 2 and Tone 3 were the easiest to make through perception. Yet both learners and NSs can produce them in speech production.

Children learning Mandarin as L1, adult NSs of Mandarin and L2 learners of Mandarin all have problems in identifying and discriminating between Tones 2 and 3. However, the causes of such failure to discriminate between them might be different. In fact the relationship between perception and production is still far from clear to confidently attribute Mandarin children’s failure to discriminate between Tones 2 and 3 as the cause for poor production. There is not enough evidence to suggest that perception must precede production.
L1 acquisition research in babbling and acquisition of the language during the first year of life suggest that production and perception may be intrinsically coordinated and more integrated in development than is usually considered. Gathercole and Baddeley (1993) proposed the concept of phonological loop in the working memory to describe this relationship. In their view, the phonological loop is a system that is specialized for the storage of verbal material. It has two subcomponents: (1) the phonological store that represents material in a phonological code that decays over time; and (2) an articulatory rehearsal process which refreshes and maintains the decaying items in the phonological store. Spoken information gains direct access to the phonological store without articulatory rehearsal but the articulatory rehearsal process maintains the spoken information in the memory. Gathercole and Baddeley (1993) indicated that although the phonological loop is present and functioning from the preschool years onwards, there is little evidence that the articulatory rehearsal process is fully operative at this stage.

In adults who are learning Mandarin as an L2, both the phonological store and the articulatory rehearsal are already fully functioning in their L1. Therefore perception and production are likely to be integrated and developed in L1. Since L1 transfer is likely to occur in the acquisition of L2 pronunciation, culture and the way learners organize the world, it is reasonable to assume that the functioning and coordinated system between perception and production is also likely to be used by L2 learners in their learning.

This sort of study has significant relevance for SEA. Many of the practices of SEA work by harnessing the "body" or "soma". The mind, after all, is not the "brain" but rather an experiencing, thinking, feeling, learning network in which perception, self-perception and production are intimately related while in no way simply reducible to each other. These
practices work by helping these networks resonate with authentic materials from the TL. In this way, SEA activates previously under-exploited capacities of the learners’ L1 machine which is more sophisticated, complex, and powerful than is yet understood.

The human mind and its embodiment remain more mysterious and unknown to us than the impression with which studies in psychology, linguistics, or even neurology often leave us. Language and the human mind are not a well-charted terrain whose waterways are now rendered easily navigable. Linguists, psychologists, neurologists and even philosophers have made bold forays along some of the more easily recognized estuaries and returned with sketchy maps. Yet we do not know enough how (or even if) these maps fit together. This is because the maps are all too often annotated in sundry languages disguising as many inconsistencies as the truths revealed. Nevertheless, the study of language and the human mind are where we inhabit and travel; and they are where we take our own existences though the right words which may sometimes be difficult to muster. Each human being exists in this unknown as profoundly as any scientist or artist can describe the conditions of such an existence.

The research we are presently reviewing does not represent a massive sheaf of discoveries. They represent an almost minimally organized, disparate body of insights more suited to providing clues about possible future directions than reliable justifications for pedagogical decisions. The point of this literature review is not so much to show that SEA stands securely in-line with the best research in linguistics, psychology or neurology. Its aim is rather to enable SEA to be situated with respect to those bodies of knowledge. However, the various clues emerging from these studies might suggest not merely possible explanations by which to interpret the results of the experiment but possible paths of further development.
Although there is no "theory of SEA", SEA is not theoretically uninformed. The findings of Gathercole and Baddeley (1993), for example, speak directly to the effectiveness of many teaching practices within SEA, going some way to explain why these practices are more effective than teacherly behaviours such as explaining and modelling. Whether or not the "phonological store" has any "concrete existence" the model presented by Gathercole and Baddeley serves us heuristically. In SEA, the mind is as an abstract machine, given to processes of perception, recording and production which are best approached not merely through denotations but through the "body". It is through the “body” that redundancies in the networks linking perception to production are created so as to enhance the effects of rehearsal and thereby reducing the detrimental effects of decay.

3.1.3 Acquisition of tones before segments

Tone acquisition studies showed that children acquire the Mandarin tone system before the segmental system. Chao (1951/1976) reported a study on his niece’s granddaughter Canta. He reported that Canta’s tonal production was nearly complete at 28 months with only a few errors but she had more difficulty with some segmental sounds such as affricates, fricatives and the liquid /l/. Li and Thompson (1977) found that Mandarin children’s accuracy in tone production is far superior than their accuracy in the production of segments. They suggest that the smaller number of contrasts present in the tone system than in the segmental system would partially account for this result. They also argue that it is easier for children to learn to control glottal pitch than articulatory mechanisms. Yue-Hashimoto (1980) claimed that her daughter’s tone acquisition was at least three times faster than her segmental acquisition.
The acquisition of tones ahead of segment in L1 acquisition seems to be a universal phenomenon in a variety of tonal languages (Clumeck, 1980; Ioup & Tansomboon, 1987; Tuaycharoen, 1977). Clumeck (1980) reviewed studies on the comparison of development of children’s tone production and segmental production in various tone languages such as Mandarin, Cantonese, Thai, Lao, and Ga (a West African language) and concluded that in all these studies “articulatory control over tone production is mastered before the completion of segments.” Jeng (1979, 1985) even suggested that Mandarin tones are acquired during the babbling period, before a baby utters his or her first words.

Taken together, these studies suggest that tones are, in fact, formalizations of particular segmental habits but let us call them "tones". In L1 children, these tones are, it seems, grafted onto a rhythmic, musical "intonationality" - a capacity for intonation - which develops by means of regularization in response to input. It is possible that the success of many of the practices of SEA is attributable to a similar process. SEA derives, in significant part, from Verbo-tonal Method (VTM) which was developed by Professor Guberina of the University of Zagreb in 1960s so as to help deaf people speak. By presenting his "students" with loud, selectively filtered materials at the optimal frequency, Dr Guberina sought to approach the mind through the soma which enhances students’ ability to produce by enabling them to perceive. Yet this perception in itself relies upon the power of an archaic capacity for intonation, probably innate, and is experienced in the "body" rather than rationally. More akin to proprioception than syllogism, this intonational capacity, exploited in different ways by VTM and SEA, is akin to the "rhythmic" or "musical" component of language which is shaped in response to input.

Furthermore, driven by the urge to become increasingly intelligible, a child or adult learner grafts consonants. More precisely, it is this capacity, no doubt located in the brain and also in
the wider expanses of the nervous system, and experienced "in the body", which, along with a growing sense of lexical features and cultural expectations, enables the subject to respond to pitches and rhythms and thereby become part of the alien system of signs and performances (i.e. the "TL").

French philosopher, psychologist and linguist Julia Kristeva (1980) presents a name for this intonational capacity: the "semiotic chora". Chora means core, with connotations of "chorus" the archaic body from which Greek drama was elaborated. For Kristeva (1980), this semiotic chora establishes early contact with the mother and maintains that relation even in adulthood when the mother is abstracted or largely absent. We need not concern ourselves with the Freudian or "deconstructive" aspects of Kristeva's work here. At present it is more important to consider what Kristeva says about the semiotic chora. For Kristeva, the semiotic chora enables the rhythmic components of language onto which are grafted signifying elements of the mother tongue (i.e. all linguistic elements concerned with paradigmatic options, which exert their own force upon "semiotic" performance in accord with the demands of "meaning"). Thus, for Kristeva, the semiotic aspects of language are endlessly at war with the signifying aspects, occasionally finding profound marriages. Kristeva tells us of the "craft" of the poet, Mayakovsky:

Rhythm. I walk along, waving my arms and mumbling almost wordlessly, now shortening my steps so not to interrupt my mumbling, now mumbling more rapidly in time with my steps. So the rhythm is trimmed and takes shape -- and rhythm is the basis of any poetic work, resounding through the whole thing. Gradually individual words begin to ease themselves free of this dull roar. (Kristeva et al., 1980:28)
For Kristeva, Mayakovsky is attempting to tell us how, in his poetry, "semiotic" aspects of language such as intonation struggle enter into symbiosis with signifying aspects (i.e. denotation and connotation).

3.2 Tone acquisition of Mandarin in L2 environments
3.2.1 Research on the tone acquisition of Mandarin in L2 environments

Research on the acquisition of Mandarin tones in a FL or L2 environments is not abundant. The available research has been conducted in diverse fields such as Chinese linguistics, acoustic phonetics, and more recently in voice recognition work. Research in the pedagogy of teaching tones has also been diverse. Consequently, different emphases and different tools have been used in the research. It now falls to outline some research findings related to the teaching of tones in the field of learning Mandarin as a foreign or second language.

3.2.1.1 Descriptive studies of tonal errors

Dr. Chao Yuen Ren observed that: “From my experience of teaching foreigners Mandarin tones (this can be said to be several decades of experience), it is difficult to teach them tones but the difficulty does not lie in the fact that they cannot imitate the tones accurately” (Guo, 1991).

The problem, according to Chao, lies in the fact the students have difficulties controlling the opposing pitches within the tonal system. This might be due to the fact that within Mandarin tones, the concept of “level” does not indicate a particular objective value of frequency. Pitch measures the individual’s highest and lowest point in his/her voice level. Guo (1991) pointed out for beginning Mandarin learners who had little exposure to lexical tones would find it difficult to determine what their own highest or lowest pitches are (within their own voice range) as well as determining how the four tones oppose one another. Some foreigners when
speaking Mandarin tend to pronounce the beginning of the Tone 2 too low and the beginning of Tone 3 tone too high. Sometimes they cannot produce the nucleus not low enough and the tail of Tone 4 not low enough and too short. SEA addresses this problem by extending the range of students' vocalization; their perception of shifts of contour; and their control in duplicating such shifts.

3.2.1.2 What changes occur in the voices of English speakers of Mandarin when speaking Mandarin and English?

In an oscillograph study conducted at the University of Wisconsin between 1970-72, it was determined that the pitch range of Chinese speakers was 1.5 times wider than the pitch range of English speakers (Chen, 1974). Thus another fundamental difference between English learners of Mandarin and NSs of Mandarin is the difference in their voice ranges. However, since 1974, no study has been carried out to corroborate Chen’s findings. Nevertheless, the finding that Chinese speakers have a wider pitch range than English speakers is generally accepted and expected because of the tonal nature of Mandarin. But whether it is 1.5 times wider is open to question. This thesis attempts to provide some limited findings on the pitch range of English learners speaking Mandarin (not bilingual speakers of Mandarin and English) to put alongside the findings of Chen’s study. Regardless, one of the benefits of SEA might be to address the relative narrowness of L2 Mandarin vocalizations.

3.2.1.3 Causes of tonal errors

One of the primary causes of tonal errors, as many researchers concur, and as is borne out in this thesis, is L1 interference. White (1981) suggested that the tonal pronunciation errors of English learners of Mandarin are a result of mother tongue interference from English intonation.
White’s prediction of errors was a result of carrying out contrastive studies on both Mandarin and English. White noted that the first significant difference between Mandarin and English is the domain of the tones. In Mandarin the tone falls on one syllable while in English intonation may spread across any number of syllables. Second, White agreed that Mandarin tones cover a wider pitch range than English intonation. Following Chao (1972), a five-level pitch range was used with 5 indicating the highest pitch and 1 indicating the lowest pitch in White’s analysis.

Figures 3.1 to 3.4 below illustrate examples of English learners and Mandarin NSs’ production of Tone 1, 2 and two realization of Tone 3 within their own voice ranges. White (1981) predicted that for Mandarin tones, while NSs of Mandarin would use the entire frequency range from 5 down to 1, English learners of Mandarin, no matter what tones they produce, would only produce them within the narrow range of 2 to 3. Therefore there is a need to train English learners of Mandarin to widen their voice ranges.

![Figure 3.1 The realization of Tone 1 by Mandarin and English speakers](image)
Third, in English stress is perhaps the most important element of English intonation. English stress is patterns involving increases in duration, amplitude and pitch. Similar kinds of stress, in
Mandarin, are important but, as we have seen from our discussion on the relation of tone to intonation, they are not identical (in form or purpose) to tone. In Mandarin, furthermore, stress syllables are louder and longer with higher pitch (Kratochvil, 1998). Thus it is not the case that importing English patterns of intonation into Mandarin creates NS-like performances as long as tones are "correct". Rather, the tones which have their contours modified by Mandarin intonation are modified in a different way according to English intonation patterns. Simply modifying Mandarin sentences with English intonation patterns often incurs great costs in terms of intelligibility. Sometimes the interference from English is so unconscious and immediate that the mere mention that a sentence is a question causes the end of that sentence to rise. SEA addressed these difficulties in large part by enhancing students' ability to produce Mandarin stress and intonation patterns rather than English ones. This involves specific practices, as described in Section 4.3 in Chapter 4, which seem to reduce the effects of interference on intonation.

In English, while duration and amplitude are closely related to stress, studies by Fry (Fry, 1955; 1958; Lehiste, 1970; cited in Hyman, 1973) and Lieberman (1957 cited in Crystal, 1969) have shown that changes in the F0 are primary cues for detecting stress. Therefore, syllables with high pitch levels, particularly high contoured tones in Mandarin, could be perceived as stressed by English learners of Mandarin. Kratochvil (1969) proposed syllabic volume as the acoustic correlate of stress in Mandarin. He proposed that the fundamental frequency of a given syllable is not the primary determinant in the perception of stress, as it is in English, but that duration and amplitude play a more important role.
White (1981) argued that because of this difference, “the English speaker who is conditioned to hear high pitch as stress will hear the two high Mandarin tones (first and fourth) which begin at a high pitch level, as stress. Moreover, the English speaker speaking Mandarin often produces these high level tones whenever the situation indicates the use of stress” (White, 1981:33-35).

The above research on how stress might affect the learning of Mandarin by English speakers, however, assumes that perception comes before production. In other words, errors in production necessarily were caused by wrong perception. However, discussion in Section 3.1.2 suggests that the relationship of perception and production is not necessarily so straight-forward especially when it comes to the perception and production of Tone 2 and 3.

Anecdotal evidence from the experiences of the researcher (based on 15 years of teaching beginning English speakers of Mandarin) would argue that certain tonal errors would probably be caused by L2 students not knowing the fact that, in order to place stress on Mandarin words or express emotion, one needs to make the words louder without changing the pitch of the words (i.e. tones). L2 students often simply try to express stress or emotion according to intonation rules of English. For example, for an English learner of Mandarin, it is acceptable to say “nǐ hǎo” when one is in a normal frame of mind and “nǐ hǎo” when one is happy as shown in the following figure:
According to Figure 3.5, when a person is neither happy nor sad, his/her production of “nǐ hǎo” conforms to the Mandrin NS model. When an English speaking learner of Mandarin is happy, however, “nǐ hǎo” is inevitably pronounced as “nī hǎo” with “nī” produced at a much higher F₀.

Lastly, the complexity of the English and Mandarin intonation systems is most likely to be the major cause of L2 student errors in Mandarin production. This is because the patterns of intonation learned with one's mother tongue exert a determined pressure upon utterance.

White (1981) also suggested some solutions for the above problems. She suggested that firstly, so as to make students more aware of the wider pitch range of Mandarin, it might be necessary to have the English speaking students practice the pronunciation of syllables with high and low tones perhaps even exaggerating the extremes. Secondly, visual resources such as a diagram which illustrates English and Mandarin pitch ranges (and has the Mandarin tones marked) should be provided. Thirdly, it might be necessary to conduct lessons explaining the stress systems of the two languages. Fourthly, as the intonation system of English is so entrenched in
the English learners’ subconscious, the teacher should strive (presumably at "point of need") to make students aware of the differences when appropriate.

White’s solutions might be useful but they are not necessarily effective because they simply involve providing students with a list of prescriptive rules about the possible origins of student errors in order to raise their awareness. There is nothing very radical in such an approach and it cannot be expected that the results will be much improved. Pitch is simply added to matters such as tone, grammar and vocabulary as something the student must continually remain conscious. Furthermore, due to lack of empirical evidence to verify such claims made through contrastive analysis, it is hard to ascertain how widespread these interference-related errors actually do occur in students’ speech production. In response to this gap, this thesis aims to provide some empirical data on how English intonation patterns interfere with the learning of Mandarin by English speakers.

3.2.1.4 Order of difficulty

Miracle (1989) conducted a study of American students’ tone errors. All the subjects involved were NSs of American English and all had at least the equivalent of one academic year of Mandarin Chinese instruction. Their teachers comprised of one NS of Chinese from Taiwan and two from mainland China. The speech material was read out by the NSs of Mandarin who were also the teachers and used as a model or control.

The data themselves consisted of 24 words that were spoken in sentences and then extracted from the sentence environment for analysis. Only lexical items in the topic positions were analysed. This choice was made “in an effort to reduce the possible interference of English sentence intonation” (Miracle, 1989:51). The students were presented with sentences in Pinyin
romanization with the tonal diacritics and English translation. The tonal diacritics were used so that the students would not have to resort to guessing the proper tone of words that they may have forgotten. The words and sentences selected were all familiar to the students. All these measures were used to make it easier for the students and to ensure that they would not feel intimidated by the task.

The words targeted for analysis contained all of the possible tone combinations available in Mandarin in one and two syllable words. They were placed in sentences in the topic, sentence initial position and in sentence final or sentence internal position.

Miracle’s findings showed that the average fundamental frequency (F₀) range of the American English students tested speaking Mandarin did not appear to be substantially narrower compared with the native Chinese speakers. These results thus tended to contradict results obtained in Chen’s study (1974). In this study, the average F₀ range for three male NSs was 97.33 Hz with a standard deviation as 8.99 Hz. The student population, it should be noted, consisted of 4 female and 6 male students. The average F₀ range excluding the female students was 94.67 Hz with a standard deviation of 23.34 Hz. However, these findings could not be interpreted meaningfully because details of the F₀ range of Miracle’s students were not provided nor analysed in detail. This renders meaningful interpretation of data impossible and therefore, in the end, could neither support nor confirm Chen’s conclusions.

Miracle measured the students’ errors against the models provided by the same sentences spoken by 3 NSs of Mandarin rather than the traditional description of Mandarin lexical tones. Though the size of NS sample was small, the tonal contours and shapes produced by his NSs concurred with research findings obtained in Dreher and Lee’s study (Dreher & Lee, 1966). It
was warranted to use the models of tones provided by NS in real communicative contexts as the standard. These models are likely to be different from the ones prescribed by the traditional description of Mandarin tones in Chinese linguistics. In other words, according to Miracle, even NSs tend to pronounce Tone 1 as 44 rather than 55; Tone 2 as 24 rather than 35 and Tone 3 as 212 rather than 214.

Yet the number of changes in the NS model sentences was very complicated and numerous while NNS errors tended to concentrate on register errors such as starting the high level tone too low for Tone 1 and so on. Furthermore, though Miracle placed the targeted words in sentence initial, sentence final and sentence internal positions, he did not analyse the effect of different positions of the words on the production of tones. In this sense, Miracle’s experimental design was more rigorous than most. However, though he embedded the tones in sentences, by analysing only the syllables in the topic position, he essentially conducted an experiment on monosyllabic and disyllabic tonal compounds. Any intonation and stress related influence on the production of tones were ignored.

Pedagogically, the conclusions from Miracle’s experiment do not automatically suggest any specific pedagogical measures. Indeed, since acoustically NS and NNS’ $F_0$ ranges, according to Miracle's study, are practically indistinguishable, Miracle's work offers no insight into why NNS have any difficulty at all with Mandarin pronunciation, intonation, tones (or any other aspects of prosody). If we were to follow Miracle, we would say that looking for differences in NS and NNS utterances in terms of $F_0$ is fruitless. Thus like Shen, Miracle also turned to register in an attempt to account for apparent phenomena which seemed to escape objective description. In other words, Shen argues that in tonal production what the American learners of
Mandarin learn “is not tonal shape-such as level, rising, and falling-but tonal register such as high, low and mid” (Shen, 1989a:40) Here, however, this is where matters become complex which makes meaningful comment difficult. For instance, if NNS make both register and contour errors, can it be predicted that a reduction in register errors might have a positive transfer effect on contour errors? Without theorizing the relationship between the two types of errors, such research data is unlikely to yield any useful pedagogical measures. Register is not, after all, an intonation pattern **per se**, but a metalinguistics response to social situations marked by a tendency to deploy certain kinds of intonation patterns. So Miracle concludes that tonal difficulties, regardless of their sequence, might eventually disappear if the L2 learners continue to study Mandarin. Thus he realizes that his work implies that there is nothing special about tones. Like all other areas of language learning, improvement in tonal production and also in prosody and grammar, is also a function of exposure and time.

Shen (1989b) conducted a study on a group of American learners of Mandarin through auditory and acoustic analysis of gathered data. These students studied Chinese for at least one semester and they acquired the pronunciation of tones through the tonal shape approach. From the comparison of the overall error rates of the five tones, the rates of errors in the production in Tone 4 and Tone 1 were significantly higher than Tones 2, 3 and 5. From her study, she reached the same conclusion as Miracle that tonal register constituted the major error for the American learners and tone shape errors occur basically with Tone 3 as it possesses two allotones: one dipping and one rising.

Shen (1989b) attributed the source of the errors to L1 transfer. She argued that since high pitch is associated with stress in English (Lehiste, 1970; White, 1981), the high Mandarin Tones 1
and 4 may be perceived as stressed syllables. Similarly, the weakest syllables in English usually have the lowest pitch; thus low Tone 3 may be perceived as a signal of weak stress. Hence, when American learners of Mandarin hear syllables that are higher than other syllables, they immediately interpret this distinction as one of stress (Chun, 1982; Shelley, 1961). Therefore Tones 1 and 4 are more likely receptive to L1 transfer than the other tones for American learners of Chinese as an L2.

Shen (1989b) also investigated the possibility of establishing a hierarchy of degrees of difficulty in tone acquisition. Findings from Shen (1989b) showed that according to acoustic evidence, in the FL environment, the hierarchy of difficulty in acquiring tones would be Tone 4, Tone 1, Tone 2 and Tone 3. In other words, Tone 4 was the easiest to acquire and Tone 3 the most difficult. Contrary to Shen, Yue (1980), in her paper on the acquisition of Mandarin tones by L2 learners, claimed that Tone 1 was the easiest to learn followed by Tone 4, Tone 2 and Tone 3. However, her claim was based on observations on children acquiring Mandarin tones as a L1. Her observation was, however, confirmed by other studies on the acquisition of Chinese tones by children as a L1 (Li & Thompson, 1977; Tse, 1978; Yue-Hashimoto, 1980).

In her pedagogical advice for teaching tones, Shen (1989b) recognizes that even though some items are more difficult to acquire than others, priority should not be given to tones -- especially not through the use of tone diacritics. Shen thus agreed with Miracle (1989) that learner errors largely came from tone register, and proposed that teachers should let students experience physically the various level of register in their own voice until they become conscious of them.

It was hoped that in this manner learners would thus be trained to change their voices from one register to another within their own pitch range and with their own proper fundamental
frequency. Furthermore, Shen (1989b) quoting evidence from Kratochvil and Tseng (Kratochvil, 1984; Tseng, 1981), warned that a wholesale dependence on the representation of tones using the five point scale (Chao, 1930) because “the actual production of tones in the speech stream does not correspond in a one-to-one manner to the phonological prediction”. However, no concrete techniques were proposed to accomplish the task of training learners to be more aware of their own pitch range. It will be noted that SEA does not rely, in any large part, upon any of the practices Shen, quite rightly, finds unhelpful. Such practices are not relied upon in SEA. However, this does not represent any substantial agreement with the findings of Shen's study. Rather, it is simply that SEA and Shen have arrived at many of the same conclusions.

3.2.1.5 Research into the pedagogy of teaching tones

The design of SEA has, on the other hand, benefited from many insights gained from Lin’s work on the pedagogy of teaching tones (Lin, 1985). Lin (1985) identified two major sources of difficulties for English learners of Mandarin tones: (1) interference from the pitch pattern of English, and over-generalization of the rules of tones as the learner progresses in the TL; (2) inaccuracy in determining, describing, and explaining the Tone 3 of Mandarin, and subtle relative differences of the tones to each other “ (Lin, 1985). Lin (1985) agreed with White (1981) that L1 interference is most likely to be the cause of many tonal errors in Mandarin by English learners of Mandarin. In terms of solving the problem with Tone 3, Lin proposed that Tone 3 should be taught as mainly low dipping rather than low-dipping-rising. This method of explanation certainly seemed to improve the learners perception of Tone 3 Lin studied (Lin, 1985). Hsieh (1996) took up Lin’s suggestion of teaching Tone 3 as a half Tone 3 rather than
the full Tone 3. Her results indicated that the confusion between Tone 2 and Tone 3 was not a
dominant factor and this confirms the efficacy of Lin’s alternative teaching strategy with
regards to Tone 3.

Lin was also the first person to critique how classroom teachers and applied linguists had
consistently followed the models of tonal analysis proposed by theoretical linguists. Lin argued
that applying models from theoretical linguistics to the pedagogy of teaching tones is deeply
flawed because, firstly, the differences in the citation forms of the tones are so subtle (especially
with Tone 3) that not only do NSs of English fail to recognize them, NSs of Mandarin often fail
to distinguish them too (Howie, 1976). Secondly, the description of the Tone 3 in citation form
is explained as low dipping-rising. However, NSs speaking at normal speed, usually pronounce
it as low dipping without the rise. Thirdly, in correcting learners many teachers tend to slow
down their speed. This tends to produce "unnatural" tones (e.g. in Tone 3, an emphatic contour
as low-dipping-rising). This is particularly the case when they are trying to correct the students’
errors. But when they return to normal speed, Tone 3 is in the most cases produced as low-
dipping. This practice tends to confirm learners in the belief of their own tone-deafness.

Woo first put forward the novel idea of using musical symbols to represent and teach tones in
1976 (Woo, 1976). In his study, Lin also put forward a set of teaching strategies including the
use of musical scales that were innovative at that time. For instance, Lin proposed that music
scales should be used to represent teaching materials or tone drills rather than using an abstract
contour line to represent the tones. In Table 3.1, a syllable with primary stress is written in
upper case letters while the one with secondary stress is in bold-faced letters. A prolonged short
vowel is attached by an arrow. These devices of representing the sounds provide visual images which are closer to phonetic realities than those used in the past.

According to Lin, the musical scales were used to represent the sounds through visual images. Thus, in addition to the auditory signals provided by the instructor and the tape, musical scales offered to visually represent acoustic features. Furthermore, when these two images from different senses reinforced each other, it was possible to expect learners to produce better tones than they could without these visual images. Visual support has since been extensively researched in Applied Linguistics. A review of research data on the use of visual support to enhance language learning will be conducted in Section 3.10 of this chapter. Many of the practices suggested by Lin have been adopted in SEA. The use of musical notion, however, is not adopted yet remains impressive in that Lin thereby demonstrates the depth of her understanding of the acquisition of intonation as intimately linked to developing musical intelligence. The connection of intonation to music had also been alluded to in Chao’s pioneering work on Mandarin pedagogy (Chao, 1972).

Lin also suggested that in teaching the drilling of both individual sounds and single syllables should be avoided. This is because sounds rarely occur in isolation in real conversations. When they occur in context, the tones usually undergo secondary modifications as they interact with other prominence-related features of Mandarin. Therefore, the teaching materials selected should contain utterances with natural tone patterns or intonation such as those used for actual communication. This suggestion is consistent with the practices of SEA.
Table 3.1 Sample of a conversation represented in musical scales

Pour syllable utterances with the second stress on the second syllable and the first stress on the fourth syllable.

DeFrancis’ (1963) dialogues for memorization can be rewritten as follows:

Baì: Göô Xìàn, nì hào ma?
Baì: How are you, Mr. Gao?
Gào: Wèi nǐ hǎo, nì shì?
Gào: I’m very well, How about yourself?
Baì: Hǎo, goìng ma, Gào Tiànlái Gào míngjiào yè hǎo ma?
Baì: Very well, thank you. Are Mrs. Gao and Miss Gao also fine?
Gào: Tiánlái, Gào Xiānghèng.
Gào: They’re all fine, thank.
Gào: Zài jiàn, zài jiàn.
Gào: See you again.
Gào: Zài jiàn, zài jiàn.
Gào: See you again.
Provision of meaningful sentences is important because both psycholinguistic literature and language teachers’ reports suggest that few learners can tolerate imitating meaningless sounds endlessly. Undoubtedly, meaningful study is more effective than mindless mimicry (Brown, 1995). However, Lin suggested that in class, while meaning of the vocabulary items should be provided, time should not be spent discussing the meanings in the learner’s mother tongue. Presenting authentic materials (rather than the artificial sentences created in textbooks) is necessary, especially in listening exercises, so that students are assisted in their task of dealing with social interactions in the target culture, developing communicative strategies, or even simply interpreting what they hear. Authentic materials are not used in SEA simply for those reasons but rather because they contain NS intonation patterns and tones.

To facilitate the imprinting of the phonetic realities of Mandarin on the learners, Lin further suggested the following steps in teaching: the teacher says each item twice while the students listen carefully. When the teacher says an item the third time, the student repeat. The instructor can give an even clearer image of the tone pattern by pointing out the contour of the tones drawn on the cardboard or overhead transparencies. The students first repeat in unison, then individually. Ideally, a model is provided each time before the learners repeat.

Unfortunately, while SEA and Lin (1985) are in agreement on many matters, no full-scale implementation of Lin’s teaching strategies seems to ever have been carried out.

3.2.2 Pedagogic measures to train students’ awareness of tone registers

In proposing a pedagogic measure to train students’ awareness of tone registers, Zhao (1987) proposes a vocal-cord method to inscribe the different registers of the tones in the mind of L2 learners. He suggested that to produce Tone 1, the vocal cords should keep tensing; Tone 2, the
vocal cords at first neither tense nor lax, then tense rapidly; Tone 3, the vocal cords lax immediately after tense, and then tense again; Tone 4, the vocal cords suddenly tense, and then lax gradually. However, Shen (1989) rejected these suggestions proposed by Zhao (1987) as beyond the L2 learners’ competence because L2 learners are neither singers nor phoneticians. They are not trained to effectively control the complicated muscle movements of the vocal cords. Shen’s objection is unfounded because Mandarin NSs did not go through special training to speak Mandarin either and yet they are speaking Mandarin. This means that it must be possible to “train” students to speak Mandarin thus controlling those complicated muscle movements of the vocal cords. The training might not, however, involve artificial training of those tones but a change of body tension and proprioception might be necessary. Chen (1974) had identified the need to widen the average pitch range of English learners when speaking Mandarin. It is noteworthy that in teaching English speakers to speak Mandarin, Shen would reject attempts to develop the vocal athleticism required. While we cannot artificially and forcefully make students enlarge their pitch range, training the different muscles and the use of the body in speaking Mandarin could be a successful way of training students to enlarge their pitch range.

Of course, if Zhao’s (1987) suggestions are to be taken up, students need to develop some sort of "proprioceptive" understanding as to what it means to tense or relax their vocal apparatus - which, in practical terms means learning to create various "musical" contours. In Section 4.3 in Chapter 4, practices of SEA are explained which harness gross body movements so as to foster proprioceptive awareness of sounds and so enable even the most un-musical student to gain requisite control of their vocal apparatus.
3.2.3 Review of current research on Mandarin Prosody

Many studies into the intonation of Mandarin Chinese have been conducted (Shen, 1985; Shih, 1986; Tseng, 1981; Wu, 1982). However, it is generally recognised that designing a method to investigate Mandarin intonation is a difficult task (Kratochvil 1968:126). Therefore, analysts in this field have chosen either to base their results on impressionistic observations (Chao, 1933, 1968; Chao, 1980; Chun, 1982; Sprenger, 1965) or to conduct acoustic experiments (Chang, 1958; Ho, 1976). However, very few experiments conducted research on free conversation produced by NSs.

In Section 2.2.3 of the previous chapter, two opposing assumptions on the interaction between the tone and intonation of Mandarin are outlined. In this thesis, it is assumed that intonation-tone interaction in Mandarin involves a pitch change in both the shape and scale of a given tonal pattern but the basic features of tones assert a powerful force upon pronunciation. In other words, do tones change according to different intonation patterns yet the contour associated with a certain lexical item exerts determining pressure, the outcome thus being the effect of complicated habits of speech which have yet to be sufficiently understood by researchers. However, while SEA is a methodology where prosody is learned through interaction with materials where tones are already subject to intonation, an academic representation of the precise nature of such interactions is not a prerequisite for progress.

3.2.4 Summary of research on tone acquisition by Mandarin speaking children as L1 and by students learning Mandarin as L2

No agreement has been reached in both acquisition of Mandarin as L1 and L2 with regards to the order of difficulty of tones. However, the research hinted that the way input is provided
might contribute to the acquisition of certain tones. Nor can the source of error be traced solely to linguistic habits such as register or pitch range.

There is universal agreement, however, on the second possible source of difficulty, which is interlingual by nature and refers to interference from English prosodic features, most significantly from English intonation. The third possible source of difficulty proposed is due to the inherent properties of Mandarin such as the complicated tone sandhi system with Tone 3.

Pedagogical measures have been proposed based on the findings regarding the sources of tone errors. These measures include consciousness raising activities about the pitch range of Mandarin, stress system and possible interference of English on Mandarin (White, 1981). Other suggested measures include not teaching with diacritics; letting students experience the various level of register of their own voice physically (Shen, 1985); using musical scales to offer visual feedback to students (Lin, 1985); not drilling single sounds and syllables but practice sentences and phrases (Lin, 1985); combating the inherent properties of Mandarin such as the complicated tone sandhi system for Tone 3 by teaching Tone 3 as low-dipping rather than as low-dipping-rising (Lin, 1985); and finally to include the body in the process of inscribing the different registers in the minds of L2 learners via the vocal-cord method (Zhao, 1987).

Though many of these proposed pedagogical measures are theoretically sound, no full-scale implementation of these measures have been carried out. SEA has benefited a great deal from these research findings which in turn have provided vital clues for the construction of SEA.
3.3 Child-Adult Differences in Language Acquisition

While children appear to master the Mandarin tone system with much greater ease and speed than the segmental system, this is not the case with foreign adult learners. It is well known that for adult learners of Mandarin, the phonological segments do not present any particular difficulty, but that acquisition of the tone system is a major task. Indeed, Lin (1985:31) claims that “few adult English-speaking learners of Mandarin have learned even near native control of tones”. Adult learners experience difficulty with other prosodic features of a L2 phonology besides tone. Many teachers and researchers have observed that adult learners learning other L2s seldom achieve native-like pronunciation, although gifted learners do attain NS mastery of other aspects of a L2.

The correlation between age and pronunciation ability has intrigued many language researchers. A considerable amount of research has been carried out in order to verify this relationship. The results have been mixed. To date, no simple, straightforward link has been demonstrated between age and pronunciation ability. Even among researchers who agree that age-related differences do exist, there is disagreement as to the explanation for such differences. A number of theories have been proposed to account for the disparity between child and adult language learning performance. In the rest of the chapter, some of the related theories are reviewed.

3.3.1 The Critical Period Hypothesis

3.3.1.1 What is a critical period?

The notion of a critical period (CP) has been influential in bio-behavioural research for close to a century. In essence, it proposes that there is a biologically determined, specific and “fixed” or invariant period of time during development during which an organism’s neural function is
open to effects of external experiential input (Werker & Tees, 2005). In its classic conception, while there are constraints on the nature of the environmental input that can modify or influence the organism, the beginning, end, and the length of this window of time are invariant. It was assumed that prior to, and after, this “critical” point in development, the system could not be altered by experience or external stimuli.

Decades of research have confirmed that in some instances, particular systems do indeed show clear evidence of critical developmental time windows, with relatively abrupt onsets and rather complete closures. However, many other biobehavioural systems have shown elasticity in either (or both) the onset and offset (Bateson, 1979; Bateson & Hinde, 1987). Such evidence has been so overwhelming that many researchers substituted the term “CP” with the term “sensitive period” (SP) (Bateson, 1979; Michel & Tyler, 2005). Indeed, in the field of language acquisition, “CP” and “sensitive period” have, by and large, been used interchangeably, thus clouding the important distinctions between the two. Therefore, to emphasize the distinction and to ensure that SP refers to a window that is more variable in onset and offset than the classic CP, in this thesis, the term “optimal period” (OP) would be employed. An OP, like an SP, differs from a CP in assuming that neither the onset nor the offset of the period is “absolute”. Tees (2001) described an optimal period as a biologically (and experientially) determined period during which some aspect of an organism’s neural and behavioural functioning is especially sensitive to a particular environmental factor. Environmentally induced modifiability outside the “best” time period is acknowledged as a real and testable possibility. Additional possibilities involve shifts in the beginning and the end of the best period depending on the organism’s stimulation history prior to and during the optimum time period (Cancedda et al., 2004 cited in Werker & Tees, 2005).
3.3.1.2 Critical Periods in Language Acquisition

Notions of a CP for SLA (SLA) have been with us for some time now. Lenneberg (1967), for example, posited that there was a neurologically based CP, ending around the onset of puberty, beyond which complete mastery of a language, first or second is no longer possible. Loss of plasticity was suggested as the primary cause for such incomplete attainment. Variations on Lenneberg’s hypothesis have been proposed by (Scovel, 1969, 1988; Seliger, 1978; Walsh & Diller, 1981).

The CP hypothesis suggests that during the CP language acquisition takes place naturally and easily, and the brain retains its plasticity. Penfield and Lamar (1959) state that the optimum age for language acquisition falls within the first ten years of life. According to them, with the onset of puberty cerebral plasticity begins to disappear owing to the inevitable and irreversible lateralization of the language function in the left hemisphere. The development of the left hemisphere of the brain was seen as the language-processing centre. There is disagreement about when lateralization is complete. Lenneberg (1967) postulates that the CP for language acquisition terminates at puberty. He infers that language learning can take place, at least in the right hemisphere, only between the ages of two and about thirteen. Krashen (1973), on the other hand, argues that language lateralization is complete by age five and therefore cannot end at puberty. It therefore appears that while the existence of the CP in language acquisition is generally accepted, its neuro-physiological cause, if any, is still unclear.

In terms of phonology, it has been argued that “it is likely that there are different OPs for the acquisition not only of phonology in comparison to other subsystems of language but for the different realms within phonology as well” (Werker & Tees, 2005:236). Scovel (1988) argued
that there is a CP for the acquisition of the pronunciation of a L2 only because pronunciation is
“the only aspect of language performance that has a neuromuscular basis that requires
“neuromotor involvement” and has a “physical reality” (Scovel, 1988:101).

3.3.1.3 Evidence of perceptual sensitivities not being lost

Though it is generally accepted that modification of bio-behavioural systems is more difficult
outside the OP, it is uncertain whether this is because the unused neurons are lost. An
experiment involving language training seems to point to the fact that the unused neurons are
not lost and might be able to be reactivated in future learning. Neufeld (1977, 1978, 1979)
conducted a laboratory experiment in which he trained 20 English-speaking Canadian
university students in the pronunciation of Chinese and Japanese, using an 18-hour intensive
perceptual training for 12 hours. The subjects made a start by imitating short phrases they heard
through their earphones. In the final phase of the program they repeated 10 short phrases five
times. Their last attempts at imitating each phrase were recorded and later played to three NSs
of each language who rated them for accent. Of the 20 subjects, 1 subject received native ratings
in both Japanese and Chinese and 2 more subjects in Japanese only. Neufeld (1977) interpreted
these findings as indicating that native-like mastery of the sound patterns of a FL is attainable
by adult learners and that L2 learners can still learn the non-native contrasts present in a FL.

However, as Scovel (1988:159) and Long (1993) have pointed out, Neufeld’s study contained
some flaws that cast doubt on the validity of his conclusions. For instance, the outcome of his
experiment may have been influenced by the instructions to the judges. The judges were not
told that the phrases they were to rate had been pronounced by English-speaking learners of
Japanese and Chinese. They were made to believe that these samples were from Japanese and
Chinese immigrants whose pronunciation of their mother tongue might show traces of interference from English. Furthermore, the relevance of the experiments would seem to be limited by the fact that the subjects were not informed about the meaning or the grammatical structure of the phrases they were trained to imitate. Therefore, it could be argued that Neufeld’s study only showed that adults can be trained to imitate or mimic phrases from a FL (which do not carry any meaning for them in a native-like manner). For his results to be accepted as counter evidence to the notion of a CP for accent, they would have to be replicated in studies using pronunciation tasks that require language processing at a deeper level than mere imitation. What Neufeld’s (1977) experiment does suggest, however, is that the original perceptual and motor abilities that allow children to acquire the speech sounds of their L1 are still available and can be accessed by adult L2 learners.

Neufeld’s research findings are further supported by Tees & Werker’s studies that found that declines in performance on non-native contrasts were not absolute and that they showed that a latent sensitivity to non-native distinctions continues to exist (e.g., Tees & Werker, 1984; Werker & Logan, 1985; Werker & Tees, 1984 cited in Werker & Tees, 2005). There are now a number of event related potential (ERP) studies which reveal neural responses to both native and non-native distinctions. However, the ERP to the non-native contrast may be slower and/or be over different recording sites than is the ERP to native phonetic distinctions. This led Werker & Tees (2005) to argue that maintenance of sensitivity to a speech contrast should be conceptualized as resulting in reorganization rather than loss.

Werker and Tees (2005) also pointed out that results found in behavioural and imaging studies of speakers of tone languages in comparison to speakers of non-tone languages suggest that
experience in learning a tone language is necessary for categorical perception of tone distinctions (Werker & Tees, 2005).

3.3.1.4 Plasticity of the brain

Tees & Werker’s research finding that a latent sensitivity to non-native distinctions continues to exist is also supported by a number of studies in language acquisition. For instance, Flege (1995) postulates that “the mechanisms and processes used in learning the [native language] sound system… remain intact over the life span” and that these mechanisms are continually at work in second-language learning. Based on these principles, Flege (cited in McCandliss et al and, & McClelland, 2002) proposed the speech-learning model (SLM) that focuses on the extent of learning to discriminate non-native phonetic contrasts. In SLM, experience still shapes the phonetic category landscape at all ages, but greater entrenchment carries with it greater resistance to change. For instance, category formation for a non-native language sound may be blocked by the mechanism of “equivalence classification”- that is, if both members of a non-native contrastive pair are perceptually equivalent with respect to the native phonetic categories, new categories may never form (Best, McRoberts, & Sithole, 1988; Werker, 1995; Werker & Logan, 1985; Werker & Tees, 1983).

Flege’s SLM model provides an explanation for why foreign accents have been routinely attested for late L2 learners. Based on experimental evidence (cited in McCandliss et al., 2002), it could be argued that access to original perceptual abilities merely becomes more difficult in some way after a certain age but not totally lost.
3.3.1.5 The effect of training after the optimum period

However, Flege’s SLM model does not propose any actual mechanism through which experience with non-native sounds leads to change. A proposed mechanism came from speech-perception training studies which focused on just what type of training leads to most improvement. McCandliss et al (2002:91) suggest that “the neural substrate of a phonetic percept may be a pattern of neural activity within some region or regions of the cerebral cortex. Given this … experience with a language may result in a situation in which a recurring categorical pattern of neural activity is elicited by a range of acoustically distinct inputs - in some cases, failing to capture aspects of speech input that distinguish non-native phonemes”.

McClelland et al (1999) carried out computer simulations of a model based on what is known as Hebbian learning in connectionism and neurophysiology. Hebbian learning is the contribution of a Canadian neuropsychologist, Donald O. Hebb. In his book, The Organization of Behaviour, Hebb (cited in McCandliss et al., 2002:62, their italics) presented a theory of behaviour based as much as possible on the physiology of the nervous system. Hebb reduced the types of physiological evidence into two main categories: (1) the existence and properties of continuous cerebral activity, and (2) the nature of synaptic transmission in the central nervous system. Hebb combined these two principles to develop a theory of how learning occurs within an organism. He proposed that repeated stimulation of specific receptors leads slowly to the formation of “cell-assemblies” which can act as a closed system after stimulation has ceased. This continuous cerebral activity serves not only as a prolonged time for structural changes to occur during learning, but also as the simplest instance of a representative process (i.e. images or ideas).
The most important concept to emerge from Hebb’s work was his formal statement (known as Hebb’s postulate) of how learning could occur. According to this postulate, learning was based on the modification of synaptic connections between neurons. Specifically,

> When an axon of cell A is near enough to excite a cell B and repeatedly or persistently takes part in firing it, some growth process or metabolic change takes place in one or both cells such that A’s efficiency, as one of the cells firing B, is increased. (McCandliss et al., 2002)

The principles underlying this statement have become known as Hebbian Learning. From a neurophysiological perspective, Hebbian learning can be described as a time-dependent, local, highly interactive mechanism that increases synaptic efficacy as a function of pre- and post-synaptic activity.

McClelland et al.’s computer simulations of how Japanese adults who have resided in the United States demonstrated great difficulty in discriminating between /r/ and /l/ showed that learning may rely on a Hebbian process, at least partly. It appeared that competition for space in representations may also be relevant to understanding cases in which learning fails. McCandliss et al (2002:91) carried out a further experiment which considered the consequences of these ideas for retraining language perception. Again, the learning being studied involved Japanese adults learning the [r]-[l] distinction. The hypothesis being tested in this study was that “it should be possible for Japanese adults to learn the [r]-[l] distinction, if only we [McCandliss et al] can find contrasting [r]-and [l]-like stimuli that will elicit distinct perceptual representations”. In this case, Hebbian learning would reinforce the tendency of these inputs to elicit such distinct representations, potentially leading to rapid progress in perceptual learning. One way to do this is to create [r] and [l] stimuli based on natural speech, in which the portion of the speech signal corresponding to the [r] and the [l] phonemes and adjacent transitions has
been acoustically altered to exaggerate the differences between them. “By using highly exaggerated stimuli that most of our Japanese native subjects do hear as different, we expect Hebbian learning to reinforce the resulting distinct percepts” (McCandliss et al., 2002).

McCandliss et al (1977, 1978, 1979) also tested the prediction under the Hebbian account of failed learning carried out by McClelland et al (1999) that perceptual learning does not depend on feedback about the accuracy of one’s perceptions. They further hypothesised that “the use of stimuli that exaggerate the contrast between [r] and [l] will allow non-NSs to learn to discriminate these stimuli, even in the absence of any feedback.” Two contrasting conditions were created. One was called the adaptive condition and the other the fixed training conditions. In the adaptive conditions, subjects were trained with initially exaggerated stimuli which were adjusted throughout the course of learning. Success in identifying the first phoneme of the word in eight stimuli in a row resulted in replacement of one of the training stimuli with a slightly less exaggerated stimulus, moving one small step inward along the [r]-[l] continuum. A single incorrect response resulted in replacing one of the stimuli with the item one step further from the point at which perception of NSs crossed over from [l] to [r]. In the fixed condition, a second group of subjects received the same number of training trials and had the same task of attempting to identify the stimuli. However the stimuli were a fixed pair of contrasting speech tokens that was reliably identified by English speakers but that the Japanese subjects could not discriminate with accuracy greater than 70%.

To assess the role of feedback, two additional groups of subjects were run, using the same conditions described above. The above groups received no feedback and the two additional
groups received visual feedback immediately after each identification response during the training phase of the experiment.

Results of this experiment were illuminating. The substantial learning in just three 20-min sessions seen in three of the experiment’s training conditions underscores the point that adult language learners maintain considerable plasticity in their ability to learn perceptual speech contrasts.

Subjects in the adaptive training condition demonstrated substantial gains in identifying stimuli on the continuum used in training even without feedback. But their boundary appeared to be in a different place than that of native English speakers. Larger gain was obtained from training with fixed training group with feedback than without it.

Training with feedback led to clear gains in discrimination between [r] and [l] stimuli for both the fixed and the adaptive regime. The evidence presented is quite clear in establishing that availability of feedback can have a dramatic effect on learning.

The Hebbian account predicts that subjects who received exposure to stimuli that they can discriminate should benefit from that exposure, even if they received no feedback on their accuracy in identifying these stimuli. In this experiment, it was found that the group receiving adaptive training without feedback, in which initially exaggerated stimuli were used, showed considerable gains in both identification and discrimination, when compared to a no training control and when compared to the group receiving fixed training with difficult stimuli.

The Hebbian account could not explain the result of why fixed training with feedback can lead to learning that is at least as good as or better than adaptive training under some conditions. It
might be suggested that both exaggeration and feedback facilitates learning by calling subject’s attention to the cues that distinguish the training stimuli. Both exaggeration and feedback would be operating to produce the same effect. One can easily see how exaggeration could help orient attention, since exaggeration would make the differences between the stimuli highly salient. One might suggest that the effect of exaggeration is to increase the likelihood that the subject will be able to generate consistent labels even in the absence of feedback.

In this experiment, only a single contrasting stimulus pair such as “rock/lock” spoken by a single individual was used. It was found that when the networks were exposed to just one pair of stimuli, their representations eventually separate, even if initially they are highly overlapping. This separation does not occur for such pairs if the set of training stimuli includes several additional stimuli. This difference in the response of the model in these two situations reflects the fact that the competing stimuli divide up the units available for representing speech sounds into clusters. However, when there are only two stimuli, they will eventually divide up the space between them even if they are highly overlapping. Slight differences in their representations eventually become amplified, leading to the gradual emergence of distinct representations.

While results from both Neufeld (Neufeld, 1977, 1978, 1979) and McCandliss et al’s studies (McCandliss et al., 2002; McClelland et al., 1999) were based on simplified experimental conditions which did not mirror real language learning conditions, these studies do offer us some directions in designing the learning environments to correct some of the errors made by L2 students. Neufeld’s (1977) findings strongly suggest that a step in the learning process which involves imitating the rhythm or intonation of the L2 learned would be beneficial. McCandliss et al’s (2002) findings provide strong support for training students with exaggerated stimuli
with feedback, plus the fact that in training students, it might be necessary to limit the training to smaller chunks or one chunk at a time rather than a large number of contrasting pairs. Several caveats need to be considered when using findings from McCandliss et al’s findings:

Any contrasts would appear in different contexts, in different positions within a sentence and with different speakers. Such diversity distribution of even one pair of contrasts within a language might contribute to the difficulty of L2 learners in acquiring particular contrasts from natural experiences.

This experiment (McCandliss et al, 2002) dealt mainly with perceptual training of a contrast pair in English. The effect of such perceptual training on production was not investigated. Therefore it might be premature to assume that this group of subjects would naturally produce the [r] and [l] contrast better in English as a result.

This experiment (McCandliss et al, 2002) dealt with the training of the discrimination of [r] and [l] in English by Japanese speakers on a segmental level. While a large number of minimal pairs such as this have been researched upon in the last 30 years, little research has been done in similar manner on the suprasegmental level.

It would be premature to assume that adopting the findings from both Neufeld and McCandliss et al’s studies would achieve instant results. While they offer us some insights about how neurophysiology might contribute to the L2 language learning process, the designs of these experiments can only offer us information about how particular steps in the L2 learning process could be organized. Neither group of investigators would suggest that the way they carried out their studies should somehow represent the entire L2 learning process.
3.3.2 The Less-Is-More Principle

Newport (cited in Jusczyk, 1998) first advanced the idea that “less is more” in her discussions of CPs in language acquisition. Working with learners learning the morphology of verbs in American Sign Language (ASL), she noticed that late learners lack the facility to learn the morphology of verbs in ASL. She suggested that limitations on information-processing capacities at a younger age could actually be beneficial for picking up the relevant cues. She noted that early learners make componential errors in which structures are produced in part, with whole morphemes omitted. By comparison, at the same stage of learning, late learners produce “frozen” structures in which whole-word unanalysed signs are produced.

Newport (cited in Jusczyk, 1998) suggests that these differences in approach appear to derive from how linguistic input is perceived and stored by these two groups. The reduced storage capacities of the younger learners may, in fact, mean that they are only able to perceive and store component parts of complex linguistic stimuli as opposed to the whole complex stimulus. Hence, for those aspects of learning that require componential analysis, their attention to such features in the input may facilitate working out of the mapping relations more effectively for early learners than for late learners. On the other hand, a larger memory capacity results in a large increase in the number of alternative mappings from forms to meanings that must be considered. As a result, older learners may find working out the mapping relations harder.

Newport (cited in Jusczyk, 1998) further suggests that using a smaller window for processing information may help in perceptually highlighting relevant units in the input.

Newport’s (cited in Jusczyk, 1998) focus was on the morphology of words. However, similar reasoning can be used to explain why infants are more sensitive to the distributional properties
of phonotactic patterns in the input as reported by recent investigations (Bley-Vroman, 1989; Ioup & Tansomboon, 1987; Schachter, 1996:159).

In short, younger learners could actually benefit from having fewer resources available for analysing information in the speech signal. A smaller processing window might actually increase the chances of finding the kind of information needed to tailor speech perception capacities to respond to the critical features of native-language sound pattern.

According to this theory, when comparing the process of L1 and SLA, our L2 adult learners, with their developed memory and attentional capacities, would actually be able to consider more of the different kinds of information in processing the L2 input rather than just picking out the relevant units in the input. So the observed phenomenon of older L2 learners not learning as well as younger learners might not be due to the plasticity of the brain. It might be due to L2 learners’ developed memory and attentional capacities in their L1. In other words, too many kind of information in both L1 and L2 are competing for memory and attention. This makes selecting the relevant units in the speech input difficult. It follows then our L2 adult learners might require assistance in highlighting the relevant units in the speech input.

In summary, with respect to whether the critical age hypothesis is valid, this thesis contends that if it does apply, it would only be in the weak form. This thesis contends that the brain can still be changed at any age even after puberty, as illustrated by both Neufeld (1977) and McCandliss et al’s (2002) studies and many other experimental studies on the plasticity of the brain from the field of neurosciences (Shepherd, G. M., 1994). Support for the use of exaggeration and feedback in training the brain to accept new data obtained from McCandliss et al’s (2002)
studies shed some light on the kind of conditions that might be necessary in inducing changes in the brain.

In terms of SLA, it might be more worthwhile to test Newport’s (cited in Jusczyk, 1998) “less is more” principle and investigate what kind of assistance our L2 learners might need in order to better perceive relevant units in the TL. The learning environment under investigation in this thesis contains the use of exaggerated movements and gestures and feedback in the teaching of Mandarin pronunciation. It is hoped that such investigation would offer language educators some information on how exaggeration and feedback can be used to overcome the differences between English and Mandarin on the suprasegmental level and what measures can be implemented to highlight the importance of tones and prosody to our L2 Mandarin learners.

3.3.3 Cognitive Explanations

Some recent proposals such as the Fundamental Difference (FD) hypothesis proposed by Bley-Vroman (1989) assumes that first and SLA are two fundamentally different processes (e.g. de Graaff, 1997). Their argument is that in L1 acquisition, children have access to the language acquisition device (LAD). Bley-Vroman contends that empirically the processes of first and SLA are marked by fundamental differences which indicate that they are essentially two different processes. Indeed, a comparison between the two shows that adult SLA manifests very few of the remarkable properties that are characteristic of L1 acquisition. In SLA there is very little uniformity; individual cognitive ability, motivation, social status, etc. play a significant role; learning involves a serious effort; L2 learners are not “equi-potential” (Schachter, 1996) for any natural language. In other words, SL learning is simply one manifestation of general cognition and is therefore no different from any other type of general learning. The FD
hypothesis also posits a set of domain-general cognitive procedures for L2 learning in the place of the domain-specific LAD operating in L1 acquisition.

Learners find languages which are typologically closer to their L1 easier to learn, and generally achieve higher levels of proficiency in such L2s. Furthermore, the learner’s L1 has been found to exert substantial influence on both L2 competence and performance. Results from recent experimental studies (Ioup & Tansomboon, 1987:343-344) suggest that SLA is sensitive to instruction and correction. However, very few L2 learners manage to acquire complete competence in the L2. Similarly, prior learning experience contributes to success. Furthermore, there is an age-related decline in cognitive ability which might result in older learners not achieving as high as younger learners in ultimate attainment of the L2.

3.3.3.1 Left brain or right brain

A comparative study of the abilities of children and adults in acquiring Thai was conducted by Ioup and Tansomboon (1987). Ioup and Tansomboon used eight subjects in their study. These were divided into four groups of two. Group I consisted of two children who were acquiring Thai as a L1; Group II were two children studying Thai as a L2; Group III were two adult beginning learners of Thai; and Group IV were two adult advanced learners of Thai who had lived in Thailand for many years. The authors reported that native-like production of tone appeared very early in the language of the children, regardless of whether the language was their first or second, but that it eluded even the most proficient adult learners of Thai.

Ioup and Tansomboon (1987) also found that for the adults, tone was more difficult to acquire than segmental features while the reverse was true for the children. In attempting to interpret their results, they found the prevailing hypotheses on maturational differences inadequate to
explain why certain aspects of phonology, those classified as suprasegmental, such as tone, should be among the earliest-acquired features of the language by children but remain difficult for even very proficient adult learners. They therefore propose a new cognitive account of the developmental variability found in acquisition of phonology. They maintain that adults and children use different cognitive systems, located in different areas of the brain, to process language input data. They argue that the earliest developing cognitive processes in children are associated with the minor right hemisphere, noting that those aspects of language associated with the right hemisphere, such as tones and intonation, have been observed to be the first acquired by the child. They hypothesized that: “The child acquiring linguistic tone first perceives and processes it as part of the prosodic system of language, analyzing the input data relevant to tones using gestalt or holistic cognitive strategies” (Ioup & Tansomboon, 1987:343).

They further suggested that “Goldberg and Costa (1981) have argued convincingly that when new descriptive systems are integrated into pre-existing codes, they are processed with greater facility by the left hemisphere. However, when there is no relevant descriptive system immediately available in the cognitive repertoire, the right hemisphere’s participation is most essential for initial orientation” (Ioup & Tansomboon, 1987:344) This, they claim, is the underlying reason why the adult subjects in their study could not attain a native-like mastery of tones. An alternative explanation might simply be that adults may not be processing tone at all especially when there is evidence that tonal categories needed to be specially taught (Werker & Tees, 2005). What adult learners did process were those auditory cues that their L1 deemed as salient. The most likely candidate for this is stressed syllables. The conclusion of Ioup and Tansomboon’s study implies that the attainment of native-like tone quality is possible in foreign
adult learners provided that adults have the opportunities to utilize the kind of right brain
dominant skills that children use in acquiring a L2.

So far, research reviewed in this chapter proposes that children process speech through the right
brain and L2 adult learners tend to process L2 speech through the left brain. The picture might
be less clear-cut than it is proposed. Some recent evidence using magnetic resonance imaging
showed that pitch contours associated with tones are only processed in the left hemisphere by
Chinese listeners, whereas pitch contours associated with intonation are processed
predominantly in the right hemisphere (Gandour, Wong, Hsieh, Weinzapfel, Van Lancker, &
Hutchins, 2000). This might be true for speakers and listeners of any language because the
processing of vocabulary items and syntax in any language have been found to be processed by
the left brain whereas the right hemisphere of the brain has been shown to process the melody
of the language. The magnetic resonance imaging evidence on Chinese listeners illustrated
precisely this point. Furthermore, since all languages are intonational and contain vocabulary
and syntax, it is reasonable to assume that both hemispheres of the brain are involved in
processing language.

3.3.3.2 Production Processes in SLA

In the SLA literature, it has often been assumed that output is only a sign of SLA that has
already taken place and that it does not serve any significant function in language acquisition
processes (Swain, 1985). However, such a limited view of output has been questioned since
Swain’s ‘output hypothesis’ was first proposed (Swain, 1985). The output hypothesis postulates
active roles played by output in the overall SLA processes. It was formulated essentially as a
reaction to Krashen’s claim about the major role of “comprehensible input” in SLA (Krashen,
1981). Krashen’s postulate of “comprehensible input” and the Monitor Model of i + 1 are based on research findings on Canadian immersion programs. It was found that while students had great success in many areas of the students’ language development (e.g. listening comprehension, fluency, functional abilities, confidence in using the L2), students had also been found to have problems in some aspects of the TL grammar especially in morpho-syntactic areas even after many years in these programs (Swain, 1985).

Swain (1985) argued that one of the important reasons for this was that these learners engaged in too little language production, which prevented them from going beyond a functional level of L2 proficiency. Swain argued that immersion students lack opportunities for output in two ways:

First, the students were simply not given—especially in later grades—adequate opportunities to use the TL in the classroom context. Second, they were not being ‘pushed’ in their output. That is to say, the immersion student developed, in the early grades, strategies for getting their meaning across which were adequate for the situation they find themselves in: they were understood by their teachers and peers. However, there appeared to be little social or cognitive pressure to produce language that reflected more appropriately or precisely their intended meaning: there was no push to be more comprehensible than they already are (Swain, 1985:249).

This description of the learning context for immersion students sounds extremely familiar to FL (FL) teachers. In other words, using the TL to teach in a FL classroom is no guarantee that the input is sufficient for acquisition because students not only develop strategies for getting their meaning across, they also develop strategies for interpreting the message in the TL given by the teacher.
The construct of comprehensible output posits that when learners experience communication difficulties, they would be pushed into making their output more precise, coherent, and appropriate, and this process is said to contribute to language learning. As Swain states,

> In speaking or writing, learners can ‘stretch’ their interlanguage to meet communicative goals. They might work towards solving their linguistic limitations by using their own internalized knowledge or by cueing themselves to listen for a solution in future input. Learners (as well as NSs, of course) can fake it, so to speak in comprehension, but they cannot do so in the same way in production...[T]o produce, learners need to do something; they need to create linguistic form and meaning and in so doing, discover what they can and cannot do (Swain, 1995:127).

Since the output hypothesis was first proposed, Swain refined her hypothesis and specified the following four functions of output (Swain, 1995). A brief summary is provided below:

1. Fluency function: output provides opportunities for developing automaticity in language use.

2. Hypothesis-testing function: Producing output is one way of testing one’s hypothesis about the TL. Learners can judge the comprehensibility and linguistic well-formedness of their inter-language utterances against feedback obtained from their interlocutors.

3. Metalinguistics function: output processes enable learners not only to reveal their hypotheses, but also to reflect on them using language. Reflection on language may deepen the learners’ awareness of forms, rules and form-function relationships if the context of production is communicative in nature. Such reflection does not necessarily require labels that constitute grammatical knowledge.
4. Noticing/triggering function: in producing the TL output, ‘learners may notice a gap what they want to say and what they can say, leading them to recognize what they do not know, or know only partially’. The recognition of problems may then prompt the learners to attend to the relevant information in the input, which will trigger their IL development.

3.3.3.3 Perception-behaviour Expressway (argument for the relationship between perception and production)

There has been a lot of debate about the relationship between perception and production in language learning. Historically, in the field of psychology, two positions with regards to the connection between perception and behaviour exist. The first is the behaviourists’ thesis that states that responses follow directly from perceived stimuli, or S-R bonds (e.g. Skinner, 1938; Watson, 1913). Therefore, perception directly leads to action (Gibson, 1979). These responses are not imitations of the perceived event but are stamped in responses to stimuli based on one’s past reinforcement history. The second apparently similar theoretical position is Gibson’s notion of affordances (cited in Bargh & Dijksterhuis, 2001). According to this view, environmental stimuli directly suggest the appropriate behavioural response to them. For instance seeing the grilled lobster says “eat me”.

Both the behaviourists and the Gibsonian theorist argue that behavioural tendencies are put into motion directly by perceptual activity. They argue that these tendencies are learned responses over time based on one’s history of reward and punishment with those stimuli. However, psychologist John Bargh (2001) argued against the above mentioned perspectives by reminding
us that the imitative nature of human beings make it possible for the relationship between perception and behaviour to be overlapping rather than sequential. He argued that

[It] is the human (and basic animal) tendency to act in the same way as we see others act. We will contend that this phenomenon flows directly from a fact of mental representation and organization—that perceptual and behavioural representations for the same action overlap. Thus the effect is a natural consequence of the automatic activation of the behavioural response by the perception of someone else doing the same thing. It is not necessary that the behavioural response be stamped in as a habit through reinforcement and it is not necessary for the response to be intended and strategic (Bargh & Dijksterhuis, 2001:2).

So is perception directly linked to action or mediated by understanding? Bargh (2001:3) pointed out that from an evolutionary perspective, “perceptual abilities and functions developed because we started to behave, not because we started to understand”. Milner and Goodale (1995:11) also noted: “natural selection operates on the level of overt behaviour; it cares little about how well an animal “sees” the world, but a great deal about how well the animals forages for food, avoid predators, finds mates, and moves efficiently from one place in the environment to another” (cited in Bargh & Dijksterhuis, 2001). In sum, perception is for doing.

However, while there is evidence that for non-primate animals, there is often a one-to-one relationship between a specific perceptual process and a specific form of action. The impressive synchrony of movement that shoals of fish display provides evidence to the hypothesis of a direct link between perception and action (Buylendijk, cited in Bargh & Dijksterhuis, 2001).

Perception in human does not always lead to a specific action because we are all able to look at a grilled lobster without starting to eat it. In other words, in human, perceptual processes can occur independently of specific actions. Thus it has been proposed, from an evolutionary perspective, in humans, perceptual activity is sufficient to create action but that it is sometimes
inhibited. Evidence supporting the “inhibitor” proposal was obtained from studies of people of various disorders such as aphasia, apraxia, low-rate mental deficiency, epilepsy and catatonic states. These are conditions in which the ability to control or inhibit thought and action is impaired. In a sense, rules of behaviour, the schema that we carry in our heads, meaning of the language we use to speak all represent inhibitors standing between perception and action.

3.3.4 Social-Psychological Factors

In contrast to the neurological and cognitive explanations given above, an array of socio-psychological factors which influence SLA have been proposed to account for the relatively poor performance in L2 learning by adults. Among models that have attempted to outline the factors and variables that affect L2 learning, the one posited by Schumann (1978) is the most influential in SLA.

Schumann (1986:379) posited the term acculturation which is defined as “social and psychological integration of the learner with the TL group”. In his opinion, an amalgam of social and affective variables affects TL acquisition.

3.3.4.1 Individual Factors of Acculturation

While socio-cultural factors may offer some explanation for conditions that might be amenable or impedimentary to SLA, affective factors concerning the individual’s capability for learning may even operate in spite of social conditions (Brown, 1987:100). Affective factors refer to the “emotional side of human behaviour” (Schumann, 1986) which include variables such as language and culture shock, motivation, empathy and “ego-permeability” (Guiora, Beit-
Hallahmi, B. Brannon, & Dull, 1972a), as well as self risk-taking, anxiety and extroversion/introversion all affect one another in overlapping ways.

The notion of empathy and ego-permeability draws on Guiora’s influential work on identity and self-representation. Guiora’s position is that social agents develop not only body egos (referring to the Freudian construct), but also “language egos” (Brown, 1987). During growth and socialization, actors form boundaries of self-representation around these egos; at early stages these boundaries are permeable, by adulthood they are much less plastic. The degree of permeability is considered to have an important effect on the ability to change one’s self and this includes practices to pronunciation to social behaviour. Parallel to this, ego-permeability reflects an ability to develop empathy for others (where empathy is defined as to project “one’s own personality into the personality of another in order to understand him or her better” (Acton & Walker De Felix, 1986; Gardner, 1985; Gardner & Lambert, 1972; Guiora et al, 1972b).

Speaking of a "language ego" might not be particularly helpful, since it implies that inhibition is a sort of repression (in the Freudian sense). That might or might not be so. In any case, it is not necessary to consider students' "inhibitions" in those terms in order to come to an understanding of how those inhibitions negatively impact upon students' learning. The idea that students have a sense of themselves as speakers, are concerned not to embarrass themselves, or are wishing to attain the approval of those around them, moreover, should not be taken as a purely inhibitory phenomena. Rather, such desires must be understood as factors leading students engage in L2 in the first place and as wishes to be facilitated through the construction of language learning environments. Fostering students' desire, in fact, is the truth of "motivation" and thus, in SEA, steps are taken to diminish "inhibition" not merely by helping students "forget themselves" (or
behave less self-consciously). These practices are, in themselves, methods by which students might gain new self-images and obtain the approval of others. This does not mean we need to see students' desire in Freudian terms. Yet the aim, from a Freudian point of view, would be to help students reshape the structures of libido invested in auditory apparatuses and thereby become capable of perceiving and producing a new range of sounds. The "Freudian" terminology, nevertheless, does not, on the other hand prohibit similar conclusions from being reached and does not seem to imply that psychoanalytic therapy will be necessary in order to reduce the negative aspects of this "language ego". Thus, in terms of L2 and culture learning Guiora hypothesizes that “ego-permeability” can be induced by lowering learner’s level of inhibition (Guiora, 1983). Although how this may be done is left for the reader to speculate.

Getting back to our discussion of CP theories, from empirical research conducted by these and other researchers (Mantle-Bromley, 1992:119), it has been concluded that integrative motivation, empathy and positive attitudes toward target groups, as well as lowered inhibitions, higher self-esteem and extrovert behaviour in interaction all enhance acculturation. Researchers also conclude that it is important to help learners “accept the frustration and ambiguity that is inherent in acculturation” (Klein, 1995).

Klein (1995) suggested that the level of FL proficiency a learner can ultimately attain results from the interplay of three factors, which he termed language processor, access and propensity. Klein (1995) observed that, although it is well known that audition deteriorates with age and that the complex and finely tuned motor control needed for authentic pronunciation also becomes increasingly difficult with age, there is no evidence of any drastic changes in these biological components of the language processor before well into adulthood. Klein (1995)
argued that absolute biological barriers to the accurate perception and production of a new system of speech sounds by late learners do not actually exist. According to Klein (1995), in order for late learners of a L2 to achieve authentic pronunciation in the L2, massive and continued access to L2 input is a necessary but not sufficient condition for native like attainment. This, he argued, can only be achieved if a learner has a very high level of propensity, a term Klein (1995:261) used for “the different motivation that push a learner forward … in the acquisition of a L2.” In other words, if learners have sufficient access to L2 input and if it is of vital importance to them to sound like a NS of the language they are learning, then it is possible that they will attain a native-like accent, in spite of a late start.

Peck (1978) found that children enjoy opportunities for "sound play" with peers (through which they get phonological practice). Peck (1978) further proposes that focusing on and manipulating sounds could be one reason children gain a better command of the phonology. Similarly, Schmid (1986) is not convinced that adult learners must necessarily perform worse than children in the area of phonology and suprasegmental production as maintained by some researchers in neuro-linguistic studies. She agrees with Burling (1981)that the average adult learner wants to learn as much as possible, and as quickly as possible, so as to meet his or her communication needs. He/she feels pressured to be able to utter complex meaning in the L2 in a very short time. As a consequence, insufficient attention is paid to pronunciation of the L2 and so interference from the L1 persists. Children learning a L2 have quite different experience. Their first and main concern is to sound like the other children; the content is added later. Schmid (1986) concludes that adults could learn from the way children acquire tone accents by emphasizing the social aspects of language learning.
SEA is not a teaching method that is built upon Peck's (1978) study. Yet Peck's (1978) conclusions help to explain some of the reasons why SEA might work since SEA utilises a great deal of sound play. SEA agrees with Schmid’s conclusion even more so since she suggests that adults neither lose the ability to engage in sound play and still profit from participating in such practices.

In summary, the relevance of this compendium of research on whether or not there is a CP, while many researchers now seem to feel that the impact of having passed the CP may not be as significant as often claimed, the following position has been adopted by this thesis. Theories that hold a "strong" position on CP would seem to suggest that it is very unlikely, if not impossible, that adult L2 learners will ever achieve near-native pronunciation (or any other features of the TL) thereby discouraging teachers from seeking to achieve high quality learning. On the other hand, theories that hold a weaker position on CP, while holding out some promise that L2 learners can possibly achieve noteworthy mastery of a TL accept that this will be easier for children. Even so, regardless of the accuracy of CP theories, the impact of such theories on SEA is, pragmatically speaking, marginal. For if there were a method that demonstrably showed that L2 learners could dramatically improve their pronunciation, it would be incumbent upon us to employ that approach, regardless of whether the results were quite “native-like”. It is the claim of this thesis, however, that students utilizing SEA often do achieve acceptable pronunciation according to NS standard, and do so regardless of the truth of CP theory. Finally, it must be noted that most of the research into CP has been concerned with the acculturation of learners in a L2 context. To what extent aspects of these findings might be applicable to learning L2 in a FL context?
The extreme diversity of learning styles, individual personality traits, and ways of dealing with the world possessed by individual students means that it would be fruitless to wait for the ideal recipe to come along to tell us how we should teach. A more practical suggestion is to take these factors into consideration in designing a learning environment that cater for an infinite number of different learning styles and preferences. The present thesis explicates and evaluates such a learning environment.

### 3.3.5 Pedagogical Environment

Children acquire their L1 in more natural settings while most adults acquire a L2 through formal instruction (Krashen, 1985). Therefore the teaching method, teaching materials, teacher’s behaviour, and nuances of the activity space all play an important role in adult SLA. In other words, the failure to acquire the phonology of a L2 may have less to do with our ability to acquire languages than with the way languages are taught.

#### 3.3.5.1 Affective factors

Krashen (1982) incorporates the notion of the “affective filter” as proposed by Dulay and Burt (1977). This filter controls how much input the learner comes into contact with, and how much input is converted into intake. The term “affective” is used because it deals with the factors such as the learner’s motivation, self-confidence, or anxiety state. Learners with high motivation and self-confidence and with low anxiety have low filters and so are likely to let in plenty of input. Learners with low motivation, little self-confidence, and high anxiety have high filters that allow in little input. Krashen (1982) proposed that the affective filter influences the rate of development, but it does not affect the route.
In using the term “affective filter”, Krashen (1982) did not describe what is being filtered out and what is being left in (presumably the brain). Krashen (1982) here makes suggestions similar to those made in connection with "language ego" though from the perspective of cognitive psychology rather than psychoanalysis. In any case, the "affective filter" is likely to have no more concrete existence than any "ego" structure that might be theorized. Neither is likely to be discovered through neurology. Rather, they are both abstract models aimed at explaining observable behaviours albeit ones that run the risk of mystifying the case by positing "psychological" objects as the cause of those behaviours. For SEA, the ramifications of Krashen’s (1982) theory of an affective filter are the same as those detailed in connection with the theory of a "language ego”.

3.3.5.2 Role of the teacher

Kenworthy (1987) suggests that firstly, teachers should persuade learners that good pronunciation facilitates comprehension; secondly, teachers should stress that intelligibility and communicative effectiveness will be the goal rather than native-speaking mastery; and thirdly they should demonstrate ongoing concern for the learners’ phonological progress. In SEA, rather than having the teacher promote the need to have good pronunciation by “telling” students that pronunciation is important, an environment is set up so that learners realize themselves that good pronunciation facilitates comprehension. This learning environment includes several forms of biofeedback and proprioceptive stimulation methods of a sort that L2 learners can further exploit outside classroom, in their private studies. In this sense, the teacher is still playing an active role in encouraging learners to improve their pronunciation but her role is one of a designer of environments and facilitator of learning.
3.3.5.3 A paradigm shift: from language learning through prescription to language learning through discovery

There is a difference between L1 acquisition and SLA. For a child, learning his/her L1 is part of the process of discovering his/her own physical environment. It occurs simultaneously with the development of motor skills and other sensory organs. Though branches of linguistics such as neuro-linguistics, psycholinguistics and sociolinguistics have tried to describe and analyze the way L1 is acquired, the acquisition of L1s throughout the world is largely unaffected by these processes and analyses because most children are successful in learning their L1s.

The acquisition or learning of a L2, however, cannot escape the force of tradition imposed by the field of linguistics, nor new developments in the theory of language, neuropsychology, psychology or even shifts in political sensibility. Among others, Lian (2000) has critiqued the dominance of the linguistic paradigm in language learning and teaching as follows:

This approach originates in the desire by people interested in the thing called “language” to develop the scientific study of language. This has given rise to the discipline of Linguistics which purports to perform proper objective and scientific analyses of language. It does so either by focusing on grammar, lexis and phonology or by modifying itself in various ways through the invention of new forms of linguistics such as sociolinguistics, neuro-linguistics, psycholinguistics or SLA - all of which use similar intellectual processes to analyse language. Linguistics also combines itself with other areas such as semantics and/or pragmatics in order to take account of the various communicative phenomena that it cannot explain on its own. Ultimately, though, the focus of all these forms of linguistics will be on items (such as words) or acceptable strings of items.
The method of inquiry adopted by this approach to language analysis is supposed to be fundamentally descriptive as opposed to normative as researchers want to know how language really is and not how some people feel that language ought to be. This supposedly objective emphasis on description is identified as a good thing as the process is supposed to reveal to us what makes up the constituent pieces of language and how these pieces relate to one another. The procedure for performing this study is essentially statistical. It relies on the collection of large amounts of string data from native-speaker informants. A statistical analysis of acceptable (and sometimes less-acceptable) sequences is performed and statistical models of how particular languages work are generated. (Lian, 2000)

Following Lian, when it comes to teaching a L2 to adult learners, the teaching is so constrained by what we are supposed to teach that we forget that language learning for any individual is essentially a process of discovery and not manufacture. It follows then that in order to produce adult learners who are proficient in learning a L2, a fundamental pedagogic move must be to bring the learning back to being a process of discovery by the learners themselves. Designing an environment where learning takes place through discovery rather than "manufacture" necessarily involves the individualization of the teaching materials and different conceptualization and organization of space for learning. SEA meets this challenge, creating a learning environment where it is possible for students to choose their own course, taking into account individual learning styles and preferences and thereby helping sustain students' desire and encouraging the development of strategies aimed not merely at developing communicative competence but negotiating new cultural situations.

3.3.5.3.1 Individualization

In this thesis, the act of language learning is treated as a way in which human beings make sense of the world through engaging with complex external multi-sensorial information. As such, language learning involves learning to cope with needs that are not always categorizable
or precisely understood. Learners also often have to cope and work with the unknown, the invisible, the unpredicted and the unpredictable.

Creating a learning pedagogy that is geared towards language learning through discovery rather than manufacture means that the individualistic and unknown needs of a large number of students and the other unknowns they might need to deal with need to be taken into account albeit they remain unknown or invisible. This means the individualization of the curriculum is inevitable and desirable.

Thus the theorem that concerns the subject of this thesis aim at creating learning stimulus that fosters students’ “connection and flows” (Deleuze & Guattari, 1987). In other words, the curriculum is designed (1) to facilitate students’ individual desires and wishes; (2) to allow more students to have their social needs catered so that they feel that perhaps they can position themselves in different ways in the social structure.

This means that asking a teacher or teachers to collect lists of needs from students in order to cater for those needs is not just inadequate; it does not work because students do not know on the first day of class where their desire will lead.

An example of using different types of things to solve a linguistic problem comes from the way my second daughter Peipei solved a problem with a phoneme in Mandarin when she was three years old. When she was three, I was teaching her the Chinese version of the “Happy Birthday” song through modelling. Most of the phonemes are easy to imitate except the sound [i] in “rì: day”. Instead saying “rì: day”, she kept on saying “yù2: fish”. In effect, instead of singing “I wish you a happy birthday” she was singing “I wish you a raw fish”. Since she could not read or
write, I was at a loss as to how to correct her. I tried explaining the mouth shape and continuously modelled the correct version for her to no avail. After a few minutes of struggling with this, she, herself, burst out and said “I know I would say it like how Shrek (a popular Cartoon character) would say it.” Then she rounded up her mouth and produced the correct retroflex sound “rì: day”. To this day, she still produces that sound correctly.

The anecdote described in the previous paragraph illustrates the resourcefulness of children and how they would find non-linguistic to solve problems constantly. In other words, since L2 learning is about coping with a new language and culture, adopting a rhizomatic learning structure to construct learning environment is like asking our students to practice coping with the L2 reality which is both the known and the unknown and unknowable and invisible. It is through such a structure that the students’ individual needs are catered for in the language learning environment that is the subject of this thesis.

3.3.5.3.2 Activity Space

In this thesis, space embraces a multitude of concepts. This ranges from the physical space of a classroom or their homes or libraries or wherever L2 students choose to carry out learning tasks as well as the psychological space of each student's engagement with the TL. This psychological space is an "abstract spaces" where students experience their sense of involvement, control or competence the speaker has in communicating his/her self, and where experiences and materials are brought into relations and thereby learners make their own maps in which they "deterritorialize" and "reterritorialize" at their own leisure (Deleuze & Guattari, 1987)
In such a rhizomatic learning environment, learners are also provided with a range of tools, both conceptual and physical, to enable them to “open their minds” to things in the target culture. For instance, they were encouraged to utilize whatever was learned to grasp socio-cultural practices of the TL community groups through weekly language exchange activities. The approach taken, however, is not so much reading textbooks and discussing, or being lectured by the teacher, but interacting, often in groups, with a wide range of authentic materials which can be interpreted in multiple ways and making demands on students’ powers of creativity.

By enabling learners to participate in creating a space for their learning, they are also able to stake one’s claims in terms of power-relations not only with other students or with the teacher but with curriculum and, most importantly, materials derived from the target culture. In their real communication with NSs of Mandarin, learners must become capable of co-managing interactions by asserting and defending themselves. Just because they are learning the TL as beginners does not mean that they need always to subordinate themselves to the power of the target culture.

### 3.4 Pronunciation Pedagogy

A particular approach to the teaching of pronunciation in Mandarin has long been taken as a given in the field of teaching Mandarin as a FL. This approach basically involves teaching with Pinyin and teacher modelling. This approach, while never claiming to render very good results, has become so generally accepted that the way pronunciation of Mandarin has been taught to L2 learners has rarely been criticized or even examined. The reason for why most students fail to achieve under this sort of regime is generally explained not by reference to any deficiency in the approach but by inadequacies in the students. Newer approaches to Mandarin pronunciation
teaching, however, have followed, albeit slowly, upon developments of teaching pedagogy in
the field of teaching English as a foreign or L2. Such approaches to the teaching of Mandarin
pronunciation are not very different from those used in the teaching of English as described by
Celce-Murcia, M., Brinton, D. M and Goodwin (1996). Due to lack of space in this thesis, I
would restrict the discussion of pronunciation pedagogy to the teaching of Mandarin only.

3.4.1 What do we teach? Segments or suprasegmentals? Inseparable!

Many Mandarin pronunciation instructors have felt frustrated by the ineffectiveness of on-the-spot feedback to students. After repeated corrections in class and many hours spent by learners exposed to traditional methods of tonal contrast drills, simplified articulatory explanations and aural discrimination exercises, many learners are still unable to produce some Mandarin initials, finals and tones acceptably. Learners tend to revert to their fossilized production of L2 once the model, assiduously drilled by their teacher, begins to slip from their short-term memory. This usually happens after a matter of minutes, if not seconds.

Redundancy in language is a well-known factor (Gimson, 1980) and intelligibility will often be assisted by an immediate linguistic and situational context in actual communicative situations. However, no matter what assistance the message is given by context, in Mandarin, suprasegmental deviance (combined with other some segmental errors) can seriously obscure meaning. This can be illustrated by a frequently used joke concerning a communication breakdown caused by suprasegmental deviance in Mandarin:

An American learner of Mandarin in Taiwan one night walked up to a dumpling hawker in Taipei and said “yī wān shuǐjiăo duō shǎo qián?” and got a slap from the dumpling seller. What he meant to say was “yī wān shuǐjiăo duō shǎo qián?” (“How much is a bowl of dumplings?”)
with the compound dumplings: “shuíjiăo”, both in Tone 3) but instead he effectively said “How much is it to sleep with you for one night?: yī wăn shuíjiăo duō shăo qián?”. “shuíjiăo” (both in Tone 4) means sleep and “wăn” means “night” rather than a “bowl of”. Both “night” and “a bowl of” are homophones in Mandarin.

This is but one such example. While clearly mythical, it illustrates that even with all the grammar and segmental aspects of the sentence correct, slight suprasegmental deviance in Mandarin can cause embarrassing communication breakdowns in spoken communication. In "real life" of course, the results are rarely as humorous (but can be just as serious). However, it is not suggested in this thesis that approaches to the teaching of pronunciation should focus only on the suprasegmental. Indeed, in Mandarin, as in English, segmental and suprasegmental aspects overlap and contribute to each other in many different, often important ways. The elements of pronunciation, in fact, combine to form a dynamic system in which, at least in practice, one feature cannot be isolated one from the others.

Faced with the complexity of the situation, armed with "critical age" theory (and encouraged to lassitude by students' apparent inability to learn Mandarin prosody through traditional methods), most teachers of Mandarin seem to have merely resigned themselves to the "fact" that their adult L2 students will never master enough intelligible prosody to become competent speakers. Yet, all agree, better control of the tones of Mandarin would make learners more intelligible and might even be able to solve the usual initial and vowel problems faced by the learners.

How should teachers structure their lessons in teaching pronunciation? One belief in the field of English language teaching is that pronunciation should be integrated into the general language lesson instead of being taught as a separate subject (Brazil, 1994). This is a belief that is shared
by the present researcher. Up until now we have dealt with what to teach; next we will review how tones have been taught.

3.4.2 Teaching segmentals

Much material has been developed for the teaching of segmentals and later some have been adapted to teach suprasegmentals as well. The most common method of teaching tones (but not prosody) in Mandarin is through perception exercises of tonal contrasts.

3.4.2.1 Tonal contrast pairs

The most common method of teaching pronunciation modification at the segmental level is based on minimal pair drills. This method assumes that the speaker cannot perceptively or productively distinguish minimally distinctive sounds such as the vowels in “chi” and “ci” because these distinctions do not exist in their L1, in this case, English. In the teaching of Mandarin tones, the practice of minimal pair drills has been extended to the use of tonal contrast pairs. In the case of Mandarin, unless the minimal pair of “chi” and “ci” is accompanied by characters and tones on the romanization, the meaning of these words would be unknown. Therefore “chi” and “ci” with neither characters nor tones are just meaningless sounds. Even with tones provided, without the characters, the meaning of these words are still not known. Traditionally, at the very beginning level, zero beginner students are usually provided with many meaningless oppositions. Students are supposed to begin to discriminate between the opposing sounds after listening to the oppositions. Meaning is usually not provided at all.

More importantly, there is not much in the way of suprasegmental context in these drills. So even if they master the drill, students often find that they cannot perform those segments very
effectively in conversation because there is no link between the drills they practiced and the
native-like prosody they craved. In fact, when beginning students do the discrimination
exercises presented in this form, they are basically mechanically listening to the contrasting pair
without relating it either to their existing knowledge or anything that resembles the Mandarin
language that they would hear in real life. Such exercises are unlikely to be meaningful to
beginners whose vocabulary and powers of discrimination are fairly limited.

Discrimination exercises based on such contrasting pairs may be pertinent to those who already
have a reasonable knowledge of the language (i.e. who have already partially discriminated
between the sounds in question). The oppositions would then act as strong reinforcing factors.
In addition, the teaching of pronunciation using such methods relies on the opposition between
single words i.e. once again they are isolated from normal intonation patterns. In other words,
such drills might be helpful as additions to the sort of practices utilized in SEA, but not as the
mainstay of a teaching method.

The minimal pair exercise was originally used by linguists for ear training. The purpose of the
exercise was to outline minimal phonetic distinctions in the vowel space. This is also an
effective exercise for elicitation of the student’s perceptual problem areas and for building
awareness of the sound contrasts in English. However, in a classroom context, teachers
encounter problems at the stage of giving feedback to the students on their production errors.
Teachers do not have sufficient time to focus on students individually. Nor do they have time to
analyse the transfer effect occurring between the L1 and the L2 or any of the intra-personal
sources of error that could possibly occur.
3.4.2.2 Minimal pairs and the language laboratory

Can better feedback be provided for in a language laboratory? Most language laboratory classes consist of students sitting next to the tape-recorders. They listen to a model and record an imitation of that model. They are then expected to “listen critically” and to compare the recordings of their voice with that of the model. A serious problem does exist however. Students, especially beginners, are unable to discriminate successfully between sounds. This is because the phonological system of L1 causes L2 learners to “filter out” perceptually acoustical differences of segments which are phonologically not distinctive in L1; thus preventing them from detecting auditorily these acoustical differences in L2. This is referred to as the “phonological sieve” (Trubetzkoy, 1939).

In a normal language laboratory situation, therefore, and despite the best efforts of the monitor, much of the students’ time is spent reinforcing errors. This does not mean that language laboratories are useless as teaching aids for pronunciation. Far from it, the language laboratory can be an extremely useful tool provided that good sensitization at both the discriminatory levels and articulatory levels has taken place beforehand. It can then act as a valuable resource. In this thesis, mechanisms similar to the language laboratory have been included in the learning environment along side a classroom procedure which ensures that good sensitization of the Mandarin language has taken place beforehand.

3.5 Teaching Mandarin tones

3.5.1 Traditional pedagogy

In traditional pronunciation pedagogy, pronunciation has been viewed as the sum of four components. The components are sounds, stress, rhythm, and intonation. In the case of teaching
the pronunciation of Mandarin, another indispensible component is the tonal aspect of the language. Perhaps it is due to the complexity of these components that traditionally in the classroom these components have been explicated separately and the teaching of tones is focused upon with very little attention being paid to stress and prominence. The truth is that NSs of Mandarin do not see a Mandarin syllable in its separate components. If you question a NS about a syllable, he/she will have no conception of the tonal component of the syllable nor will he/she know what the segmental feature is either. What he/she can tell you is how to say it and how it is written in character form and the meaning of the word. Yet, despite the relatedness of the components of the syllables in Mandarin, teachers still teach them separately when we teach Mandarin as a foreign or L2.

The common practice of teaching Mandarin is by presenting teaching materials with romanization such as Pīnyīn, plus Chinese characters with diacritics that indicate tones on either the characters or the romanized forms of the characters. In most cases, the teaching of correct tone production of lexical items in isolation has been seen as the most important aspect of the teaching of Mandarin pronunciation teaching (Jiang, 1996). Romanization is usually adopted because it provides a way of denoting tones within the spelling so that they are memorable for students. Debates discussing the relative merits of using Pinyin versus tonal spelling in the early 1990s (McGinnis, 1997) did not produce any convincing evidence that either method of romanization improves students’ production of tones. In fact, McGinnis’ study came to the conclusion that the use of Gwoyeu romatzyh- a romanization system in which the tones are built into the spelling of the romanization (the system used in Chao’s Mandarin Primer) actually reflected slightly lower rates of tonal production accuracy for NSs of both American English and Japanese (McGinnis, 1997).
To date, the studies on the teaching of tones assume that the starting point for accurate pronunciation is to be able to recall what the tone of a given syllable should be (Orton, 1992). This process of recalling tones is supposed to be helped by the tone diacritics provided on either the characters or the romanized form of the characters. Jiang’s study investigated the relationship of remembering tones and pronunciation by testing students’ tones in written work on a weekly basis (Jiang 1996). From a survey he conducted to gauge the effectiveness of using written test to test tones, he found that a third of the students pointed out the fact that the written tests on tones allowed them to remember which character went with which tone but did not help them in recognizing what the tones sounded like upon hearing. The written tone test was a mechanical test because it was done by hand and when hearing Chinese through the ears, the students still could not differentiate the tones. Consequently, from his experimental data, he concluded that pronunciation problems arise out of not being able to differentiate the four tones through perception. Improving pronunciation has to start from training students in the act of listening.

It seems that researchers are still unsure about how to effectively teach students tones of Mandarin. No convincing evidence has shown us that consciously memorizing tones works; nor does it show that using romanization to memorize tones produces an effect. At best, it shows that memorizing tones has an indirect effect on tone production. However, in traditional methods of teaching Mandarin pronunciation, the inclusion of romanization and tone diacritics is so entrenched that no textbook, so far, has been produced without romanization and tone diacritics. Although several researchers such as Shen (1989b) and Jiang (1996) had suggested that tone diacritics should not be used in teaching, very few teachers are brave enough to go against the trend.
### 3.5.1.1 Error Correction

Correcting the L2 students’ mispronunciation is extremely important in terms of providing feedback. But I would argue that not all forms of feedback are necessarily beneficial for learners. Correcting learners with the wrong type of feedback may not actually achieve the aim of motivating learning. This is particularly relevant in the study of Mandarin because the foreign learner’s relationship to tone is very complex and not very well understood. What English speakers need to do is concentrate on what a Chinese person concentrates on: tone as an integral part of the syllable. As Orton (1992) pointed out “When English speaking learners stress what is common to those syllables by attending only to the consonants and vowels, they are ensuring that they never reach native speech.”

The following anecdote from Orton (1992) is particularly illuminating:

Whenever I am having a conversation with a Chinese in Chinese and I say I am not sure of the exact word I want but that I think it is, for example, chī, but this turns out not to be meaningful for my Chinese friend, in such a case, if I then speak again, I will immediately suggest that if it is not chī, perhaps I mean chī, chí or chī. But if the Chinese speaks first, he or she will suggest that if it is not chī, then perhaps I mean cī, shī or even zhī. At such moments I am always amazed. How could anyone mix up chī and shī, or chī and cī?” I splutter, “Why, you’d never learn to speak if you did, everything would be nonsense!” Two seconds later it is my Chinese friend’s turn to be amazed. “How could anyone mix up chí and chí? Why, you’d never learn to speak if you did, everything would be nonsense! (Orton, 1992)

The above anecdote illustrates how the separateness of tone from the actual sound and shape of a word is usually taught in a teaching Mandarin in a FL class. In a Mandarin language class, it is quite common to have a dialogue such as:

Student: What’s that character?
Teacher: Tian

Student: (writes down t-i-a-n, then looks up at the teacher and asks) Which tone?

Teacher: First tone.

Students would learn to listen for tone themselves quite quickly and if they were uncertain, they would ask questions such as: “Is it first tone?” However, this pattern of seeking or providing feedback, nevertheless, continues to emphasize the separateness of tone from the rest of the phoneme or a word.

As we aim for our learners to integrate tone and its phoneme throughout the learning process, they might be better served if they adopt NS behaviour in NS-NS interactions when clearing up uncertainties over words in Mandarin. In native speaker to native speaker interaction, if uncertainty over a character or a word arises, the conversation would be more like this:

Native speaker A: nǐ jiào shénme míngzi? (What’s your name?)

Native speaker B: wǒ jiào Zhāng zhēn. (I am called Zhang Zhen.)

NS A: něi ge zhāng? (or shì gōngcháng zhāng ma?) (Which “Zhang”? or Is it “GōngchángsZhāng”?)

NS B: shì. (yes)

NS A: něi ge zhēn? (Which “Zhēn”?)

NS B: zhēnyán de zhēn. (“Zhēnyán’s Zhēn”)
If the compound “zhēnyán” is understood by A, then the uncertainty is cleared up. Otherwise, further clarification of similar nature is used until the uncertainty is resolved. In this NS-NS exchange, the problem is not focused on the tone but on the meaning of the characters (as there are many homophones for both “Zhāng” and “Zhēn”.

However, in a class situation, teachers often correct students’ errors not through the above method but by saying “No. It is not the third tone. It is the second tone”. This statement is usually followed by a modelling of the phoneme in question with the correct tone. The effect of such correction, at best, is a temporary correct response from the student of the same phoneme through imitation. At worst, such attempt is futile in providing any useful feedback to students. For instance, with Tone 3 NSs speaking at normal speed, usually pronounce it as low dipping before all tones except another Tone 3 and then with another Tone 3, the first Tone 3 becomes Tone 2. However, in correcting learners with Tone 3, many teachers tend to slow down their speed. This tends to produce Tone 3 unnaturally thus rendering it as low-dipping-rising rather than low dipping. This is particularly the case when they are trying to correct the students’ errors. But when they return to normal speed, Tone 3 is in most cases produced as low-dipping. This practice has quite often made the learners believe even more that they are tone deaf.

3.6 The shift to focus on the suprasegmental in teaching pronunciation, is it warranted? 3.6.1 Evidence from L1 acquisition

L1 acquisition research on the acquisition of spoken language have provided ample evidence that the acoustic structure of utterances contains a lot more important information that it was thought previously. Peters (1983) hypothesized that the salient chunks of information that the language learner pulls out of the speech stream could provide a means of prosodic
bootstrapping. Prosodic Bootstrapping or “Bootstrapping from the Signal” is the idea that learners could use information in the speech signal to gain valuable knowledge about the syntactic organization of their native language (Gleitman & Wanner, 1982). One possibility that this can happen is that elements of the syntactic organization are marked in the acoustic structure of utterances.

Since the 1980s, investigators such as Gleitman and Wanner and Peters (Gleitman & Wanner, 1982; Peters, 1983 cited in Jusczyk, 1998) began to suggest that information in the speech signal in terms of prosodic (and other) markers could provide clues to syntactic units such as clauses and phrases. They suggest that words were initially coextensive with stressed syllable thus learners are more successful in segmenting the speech wave into words on the basis of stressed syllables. Furthermore, language learners are sensitive to the information about sequences of sounds present in the speech wave. Language learners actually use this information to provide clues regarding syntactic organization. Based on empirical studies that they cited on adults and 7-year-olds (Read & Schreiber, 1982), they suggested that “an infant who is innately biased to treat intonationally circumscribed utterance segments as potential syntactic constituents would be at considerable advantage in learning the syntactic rules of his language” (Gleitman & Wanner, 1982:26)

3.6.1.1 Sensitivity to Prosodic markers in the Input

Research presented in the previous section indicated that some of the prosodic marking of syntactic units is present. However, do we know whether infants are actually sensitive to its occurrence? Hirsh-Pasek et al (Hirsh-Pasek, Nelson, W. Jusczyk, Cassidy, and, & Kennedy, 1987) hypothesised that if infants respond to prosodic markers in the input, then they should
prefer to hear speech that is segmented in accordance with these markers as opposed to speech that is segmented inappropriately. In the experiment samples of speech from a young woman who was talking to a 19-month-old child was collected. They excised a set of passages that were five to seven sentences in length, and they inserted a series of 1-second pauses in each of them. Two versions of each passage were produced. In one case, all the pauses were inserted at the boundaries between two clauses (from here on referred to as the coincident versions). In the other case, an equal number of pauses was inserted but between two words in the middle of a clause (from here on referred to as the non-coincident versions). They used the headturn preference procedure (please refer to the Glossary section) to present samples of both types to a group of 10-month-old babies. The infants listened significantly longer to the coincident than the non-coincident versions of the samples. In another experiment on a group of 7-month-olds with the same procedure, infants as young as 7 months old displayed the same listening preferences for the coincident versions. Hirsh-Pasek et al (1987) interpreted these findings as an indication that infants as young as 7 months old are sensitive to the presence of prosodic markers to clausal units.

A subsequent investigation by the same researchers examined whether the exaggerated tones of child-directed speech may have enhanced the infants’ ability to detect markers to clausal units (Kemler Nelson et al, 1989 cited in Jusczyk, 1998). The result showed again that the 81/2-month-olds who heard the child-directed speech displayed a significant preference for the coincident versions. This prompted Kemler Nelson and colleagues to conclude that the exaggerated tones of the child-directed speech samples may have facilitated the detection of markers of clausal units in these samples. For a more detailed review in this area, please refer to Jusczyk (1998).
3.6.1.2 Evidence that prosodic organization is actually used in infant’s speech processing

Given that infants demonstrate some sensitivity to potential prosodic markers of syntactic units, the next question is whether and how they use this information in organizing their representations of speech. Research findings from Mandel, Jusczyk, and Kemler Nelson showed that the prosodic packaging of clausal units seems to facilitate even very young infants’ memory for speech information (Mandel, Jusczyk, & Nelson, 1994; Mandel, Kemler-Nelson & Jusczyk, 1996).

The learning materials used in SEA are based on sentences with all the aspects of intonation preserved. If we take heed from the evidence obtained through L1 research on infants, then it is possible that adult L2 students of Mandarin would also use clausal information to segment language stream in L2 which may also result in their acquisition of Mandarin grammar. Similarly, adult L2 students would probably also find that such sentences are easier to remember.

3.7 Rhythm

Another neglected area of language pedagogy is the role played by non-verbal and paralinguistic features of speech acts in communication. When we think of communication, we often think only of expressing ourselves clearly: knowing what we want to say, choosing appropriate vocabulary and the correct structures, getting the pronunciation right and we are there. However, studies have shown that the non-verbal and paralinguistic features of the speech act may be of equal importance in communication and sometimes of even greater importance than the verbal elements.
3.7.1 What else do infants pick up from their interaction with their carers?

Research reviewed in the previous section described various findings from the field of L1 acquisition. Thus the emphasis is on how infants pull out information from a stream of speech. However, most of the research reviewed was conducted under experimental conditions. The social context in which caregiver-child interactions take place was not explored in detail. Like every other human activity, caregiver-child interactions, whether linguistic or non-linguistic, are social interactions even when children are too young to verbalise their responses in the participatory process.

The powerful force that enables infants to participate in the joint-engagement with their caregivers without language is rhythmicity. As Maier points out “Rhythmicity provides a “glue” for establishing human connection” (Maier, 1992).

Rhythmicity has been recognised as a potent ingredient in everyday life. However, only in the last few decades has rhythmicity been researched upon scientifically for its intrinsic power in human relations. Such research has spawned a whole new branch of psychology known as kinaesthetic. Paul Byers, one of the pioneers of this field even thought that they had located “the molecules of human behaviours” (cited in Kuhn, 1976:49).

This research and practice is in line with recent scientific emphasis upon non-linear perspectives (Gleick, 1987). An exploration of the forces of rhythmicity also reflects the notion of “chaos theory” with its stress upon minutiae as the key elements of energy and change (Moustakas, 1981:20). These more recent perspectives upon life steer away from a Darwinian obsession with locating and pursuing a single hierarchical superstructure for human existence and move us toward the diversity of life with each person’s multi-potentials. This sort of work,
while not the basis upon which SEA has been developed, nevertheless has raised interesting
questions for certain practices in SEA insofar as it can be demonstrated that gross motor
movements and prosody can remain inter-related in adults.

Maier further points out that new understanding of the impact of rhythmicity upon close
personal relations can contribute to better care practice. Researchers have learned from the
interplay between infant and caregiver that they had to find and fulfill each other’s rhythm, i.e.
they have to fall into “the rhythm of relationship” (Maier, 1992:4). As Piaget observed “In
other words, it is not merely the content of the phrases exchanged or the nature of the activities
in which they are engaged. Significantly, it is also the give-and-take with clear cycles of
approach and withdrawals, continuously maintaining each other’s rhythm, that create
meaningful relationship” (Piaget, 1954).

The implications of these rhythmic experiences for language learning are significant. Language
acquisition processes, be it L1 or L2, always involves language use in rhythmic prone
opportunities such as greeting someone with words and then with an embrace, handshake, or
casual wave in a kind of dance. Therefore, it stands to reason that the teaching of language used
in such social interactions should be taught in their rhythmic contexts.

3.7.2 Rhythmicity as a Source of Prediction.

In infancy, the order in the world is discovered through baby and caring adults attempting to
find each other’s rhythms (cited in Moore & Yamamoto, 1988). They both gain an awareness
that life is and can be predictable. Such predictability can be influential in furthering a sense
of well being and internal cohesion and is useful in relieving stress or frustration. Repetition is
the key of such predictability in rhythmic experiences.
3.8 Body movements

Another important technology in the rhythmic dance of language use and social interaction is body movement. Body movement has always been an indispensable part of the rhythmic dance in any culture. Highwater (cited in Moore & Yamamoto, 1988) has attributed this as a spiritual dimension to dance. He hypothesized that “through dance, primal people “touch unknown and unseen elements, which they sense in the world around them” (Highwater, J. cited in Moore & Yamamoto, 1988).

In preschool education in every culture, rhythmicity and body movement are powerful agents for learning. However, in adult life especially in the West “the idea that spiritually can be associated with the body is extremely remote from the Western belief in the dichotomy of mind and body, spirit and flesh” (Moore & Yamamoto, 1988). Divorcing spiritual values from our physical activities has had an impoverishing effect on the quality of life.

In field of study called kinemorphae, “the proprioceptive aspects of body sensations connected with gravity are kinaesthetic, spatial situational and spatial motional. The kinaesthetic sense is related to the tactile sense since the vestibulum organs in the ear are tactile hairs that provide the sensory data, spatial situational gives the spatial orientation of the head with respect to the earth’s gravitational field (up and down), spatial motional measures the acceleration”(Condon & Ogston, 1966).

In normal behaviour, speech and body motions are precisely and rhythmically coordinated: they are synchronized (Condon, 1980:58). Studies have shown that all body motions — blinking, waving to someone to cross the road, thumping the table — are synchronized with the rhythm of the speech accompanying them. Another way of saying this is to say, when we speak we
dance. Furthermore, the body of the listener dances too, in rhythm with that of the speaker (Condon & Ogston, 1966:338).

Movements are thus seen to be rhythmically integrated, co-occurring components of total speech behaviour, not frills or optional extras to be added later if there is time. Teaching students to pay attention to rhythm in oral skill production and relate their body and gesture synchronically to their own and their interlocutor’s speech rhythms is thus to teach students an essential and inseparable piece of competent language use.

An even more important reason for working on rhythm is that, from the same studies, the acquisition of a language’s rhythmic structures is recognized as fundamental to the successful acquisition of the language including its structures. Babies synchronize to adult speech rhythms within two weeks of birth; and they do this to the rhythms of any language: for example American babies will synchronize happily to spoken Chinese. But they will do this only to flows of speech, not to single sounds or tapping. Condon puts it: “infant lives itself into language and culture” (Gassin, 1990). As they develop, all babies babble before they speak and they do so in rhythm and with correct intonation patterns. Chinese babies are already babbling in tone before they speak. It is even speculated that the picking up of the rhythm of a language also enables infants to pick up the syntactic structures of their L1. The phenomenon of prosodic bootstrapping discussed in Section 3.6.1, might indeed be one of the reasons why all children acquire their L1 in the same length of time and with relative “ease”.

It can be hypothesized from this that L2 learners’ discomfort in speaking or listening may come from a subconscious knowledge that they are out of synchrony with what they are saying or what is being said to them — they don’t know the dance — and this leads to a feeling of
distance and sense of isolation. And these negative emotions push them to give up. This is probably more true than saying that problems of prosody are caused by mistakes regarding "register". By contrast, because children are more flexible in adapting to new rhythms this may be one of the reasons they are more successful at L2 learning. This is not to contest the notion that the child's brain is more "malleable", but rather to see that malleability from a behaviour standpoint. Flexibility of rhythmic performance is an aspect of rapid neural growth. Indeed exposure to new rhythms seems to stimulate such growth which is expressed behaviourally. Thus Gassin (1990) points out that these studies imply for L2 learning that the greatest difference between an L1 baby and an adult L2 learner is that the L2 learner lacks a fundamental perceptual and organisational framework of rhythm. They do not have the rhythm of the TL and must pick it up at the same time as picking up the lexical and structural content. Furthermore, they are encumbered with the obstacle of the native rhythm framework. In other words, good adult language learners are people who, like children, can pick up the new rhythms of a L2 without special help. As for those who have difficulties, we need to investigate whether they actually lose the ability to pick up the new rhythms, or whether unless it is taught to them, they simply do not know that they need to do so, or perhaps do not know how to do so.

3.8.1 Review of therapeutic use of body movement

3.8.1.1 Verbotonalism

Body movement has been used extensively in different therapeutic contexts. As far back as the 1960s, body movements in conjunction with rhythm had been used in the rehabilitation of the deaf in the work pioneered by Professor Guberina of the University of Zagreb. Professor Guberina pioneered a new therapeutic method known as the Verbotonal system of Phonetic Correction (Renard, 1975) for treating deaf children. This method differed from conventional
approach to teaching the deaf in that it concentrates on using the children’s residual hearing in learning to hear and to speak. In introducing the methodology, Professor Guberina and Carl Asp writes:

Underlying the method is the conviction that language evolved from spoken language and that speech (which is used interchangeably with spoken language) is a social event. We speak (i.e., we use spoken language) when we want to express something or when we react to an event. In this sense, the "meaning" of speech is transmitted not only by linguistic elements but also by the auditory and visual information present in the rhythm, the intonation, the loudness, the tempo, the pauses, the tension, and the gestures of the speaker. Thus the individual speaker is both a producer and a perceiver of speech. Most important, the auditory and visual information in his production reflects how he perceives speech. If his perception changes, his speech will also change. If we have corrected his speech, we have corrected his perception.

The verbo-tonal procedures follow the pattern of language development observed in babies who have normal hearing. Before a baby learns to speak, he cries, babbles, and coos – he produces sound. His whole body participates in producing and receiving sounds. This vocal activity is not a response to his sense of hearing; rather, it is a response to his proprioceptive sense.(Guberina & Asp, 1981:1)

This method has been successfully applied and adapted to other therapeutic and language learning contexts (Asp, 1973; Asp, Youngsun, Madeline Kline, & Davis, 1999; DiJohnson & Craig, 1971). DiJohnson and Craig (1971) undertook a project to investigate the use of Verbotonal method with preschool deaf children in 1971. In the project, the children in the Verbotonal group used SUVAG I to extend from 1 to 20 kilohertz and then SUVAG II is used individually to locate and utilize the child’s optimal hearing range. As Professor Guberina explained
First, we have an instrument called the SUVAG I, which can transmit frequencies as low as 0.6 cycles. This means lower than one cycle, and it transmits to an upper limit of 15,000 cycles. This instrument offers profoundly deaf person the opportunity to perceive very, very low frequencies and to perceive some of the high frequencies of speech, due to the perception of low frequencies. SUVAG I is used in diagnosing hearing loss.

The other instrument is called SUVAG II. It is an auditory training unit which contains many combinations of acoustic filters. We can select different combinations of filters to transmit speech through selected bands, thus we can select combinations which give the best speech perception to a given hard of hearing or deaf person (Guberina, 1969:5).

Teachers in the Verbotonal group used the Guberina method and the SUVAG equipment. For each child, “the speech sounds were transmitted in the most appropriate frequency range which can be determined by the band which is best suited for the transmission of a particular sound and that band which the diagnosis of the child’s hearing loss suggested as the most productive one”(DiJohnson & Craig, 1971:15). Other speech stimulation used included (1) body movement, (2) musical stimulation, (3) implementation and (4) individual work.

In the first part of each session, body movements were utilized to facilitate emission and correct articulation for the desired sound. For instance, for the sound of “p” in “pop”, the teacher worked with the children from a relaxed position to a tense position. During the period of maximum tension as when the teacher and the children are stretching, the children and the teacher spoke the word “pop”. Because the sound of “p” is a tense sound, it was always taught when the body assumed the position of maximum tension. However, the teacher taught the voice consonant “b”, which is less tense, by going from a tense position to a relaxed position.
For the second activity, musical stimulation, the children and the teacher simultaneously clapped and spoke phrases in the rhythm which is functionally best for the syllables to be learned.

For the implementation of the language building session, the teacher attempted to relate the therapy words to meaningful language concepts. During individual session, the teacher tried to get the child to respond to the same stimulation without amplification.

The results of the project suggested that in terms of connected speech, significant differences favouring the Verbotonal group were evident along the dimensions of intelligibility and intonation, strength and pitch. Other competencies such as lip-reading and social skills were also better for the Verbotonal group. One important finding of the project is that for the Verbotonal group, social competency remained a more important factor in predicting speech production than for the control group. For the control group, hearing loss was an important factor for poor speech production. In other words, this study seems to suggest that the Verbotonal method made the children more sociable which led them to be more competent socially, relegating the effect of hearing loss to a secondary place.

There are many similarities between SEA and the procedure described in Dijohnson and Craig’s project (Dijohnson & Craig, 1971). In a sense, one can liken FL learners to children who are suffering hearing loss. In FL teaching, rather than trying to teach deaf children hear and speak using their residual hearing, we are teaching our L2 learners to hear and speak a L2 when they already have acquired their L1. One additional aspect of the Verbotonal method is that the teaching is conducted in a highly interactive, relaxed and enjoyable atmosphere in which being involved in social activities is essential. The result of this is, as Professor Carl Asp pointed out
as far back as 1969 that the children tend to vocalize more after being treated with the Verbotonal method (Asp, 1972). For a detailed discussion of SEA, please refer to Section 4.3 in Chapter 4.

### 3.8.1.2 Speech-movement therapy based on sense physiology

Based on the system of sense physiology described by Steiner, Dr. Annelise Brüll’s research used speech movement to treat two boys suffering from mild dyspraxia (Brüll, 2003:61). Steiner (1996) argued that it is important to enable a child to have a direct sensory perception of speech sounds, phonemes, syllables and intonation in order for him/her to experience the sense of word. As Brüll pointed out “The sense of word is quite different from the sense of hearing, as its function is to pick out speech sounds amongst other sounds, noises and tones” (Brüll, 2003). Contrary to most theorists, Steiner (1996) did not see this as a cognitive process. He argued that the child was able to pick out speech sounds and practice them, before being able to judge their meaning.

Steiner also noticed the synchronous nature of the dialogic conversation between two speakers thus acknowledging the need to study the effect of the listener on the speaker as well. He described the sensory process in the following terms: “A speaks and B listens. When A speaks the larynx of B vibrates in exact synchrony with A’s speaking larynx. B then makes these movements of his larynx conscious through the appropriate nervous system” (cited in Brüll, 2003:60).

Steiner’s statement cited above has been vindicated by William Condon’s work in the field of linguistic kinesia. He showed that through interactional synchrony the speaking of the speaker influences directly the body movement of the listener: “When A speaks and B listens there exist
an exact and flowing reaction of movement from listener in synchrony with the speaker in 40-

In other words the voluntary muscles which allow us to move become an instrument of
reception and resonance. This is precisely what Condon is arguing when he states that the
listener’s muscles are the perceptive organs for the speaker’s muscle movements. Steiner coined
the new art of movement as “eurhythmy”. Brüll (2003) provides a rudimentary description of
eurhythmy as follows:

> Each consonant and vowel is expressed by a specific gesture of the arms. These gestures cannot be executed in a mechanical manner but need a “flowing awareness” or an experience of the inner quality of the gesture. Steiner states that these are not “thought-out” gestures, but that they correspond exactly to the air-movement brought about when a particular sound is spoken. (Brüll, 2003:61)

Oordt (cited in Brüll, 2003:61) has expressed it in this way:

> The movement that is invisible but audible in the sounding word is to be made visible in the movement of the whole body. Audible sounds transformed into visible movements, into visible speech: that is eurhythmy.

Brüll (2003) carried out experiments to test the hypothesis “Can the use of movements induce and enhance speech?” In the structured therapy sessions carried out by Brüll, she treated two boys suffering from mild dyspraxia who were unable to communicate using language. The sessions were conducted for half an hour each for two to four times a week in a well-sized room which allowed space for moving.

The materials used were simple rhythmical poems. They started with the training of gross motor movements through walking, clapping to rhythms. These steps helped to train the children to learn to stop and to differentiate between the meaning of “long” and “short”. Then marching
was used to specifically teaching the boys how to stop by putting their feet down decisively. Once these gross motor skills were mastered, finger games were used to transfer the function performed by the previous gross motor movements (such as stop) to finer motor skills (finger movements focussed particularly on tongue movements).

Then the rhythmic nature of the verses was reinforced by marching to the beat of the rhymes and then lastly the eurhythmy gestures representing a particular “letter” would be introduced. Once these gestures became familiar to the boys, they would do them each time they heard the corresponding letters in simple verses. Thus the perception of different sounds was sharpened. The boys were then asked to make the gesture for the target sound each time they heard the sound.

Brüll further notes: “It has been my experience that intensive practice of eurhythmy gestures, large and small with their arms, feet and fingers, eventually helps the child to pronounce the consonant. Even after years of practice, when the child has actually acquired the sounds but suddenly at the end of a word hesitates e.g. book is boo . . ., I only need to indicate the gesture for “K” and the child, relieved by the visual stimulus completes the word “book” (Brüll, 2003:62).

Lastly, after the movement exercises, the therapy proceeded to put the sounds practiced back in their rhythmic context.

The two boys followed the normal speech development trend of periods of stagnation and a sudden leap forward. In the fourth year of therapy, one of the boys began to use sentences and in the sixth year, he began to transfer practiced words to everyday situations outside therapy
sessions and also picked up words that had not been practised. As Brüll puts it “this showed that the motor-memory for words was beginning to work” (Brüll, 2003:62).

The therapeutic procedure described in Brüll’s (2003) experiment has much in common with the classroom procedure in SEA. Both focus on the rhythm of the language rather than the target of the treatment. In the case of Brüll’s (2003) study, the target of the treatment was the consonant difficulties in English. In the case of this thesis, it was the targeted treatment of tones and prosody. In both cases, the practice of the rhythm of the language was first carried out through marching movement, then clapping to train gross motor movement. Then gestures were introduced to train the finer motor movement. In the case of Brüll, such finer motor movements and gestures were used to remind patients the consonants they were practicing. For students learning Mandarin in SEA, finer motor movements such as a number of gestures to represent tones were taught to students to remind them of tones. In both cases, the last step was to reintegrate the gestures and movement and sounds back into the rhythmic sentences they started with.

3.9 Learner-centred classroom approaches

L2 teaching approaches have been moving away from didactic approaches to learner-centred approaches for a long time. In the learner-centred classroom, activities related to the correction of pronunciation errors are designed to meet the student’s differing learning styles, namely auditory, visual, tactile and kinaesthetic learning. Conformity to the learner-centred model does not necessarily make the method an effective tool for modification of pronunciation, but it is likely to make the lesson more enjoyable for the student. Many innovative but low tech
practices, as opposed to high-tech computer-based approaches have been proposed (Brown, 1995).

Accompanying the move to a learner-centred classroom approach to curriculum design, the debate of whether it is more effective to teach segments or suprasegmentals of a L2 resurfaced. Numerous reports advocated one approach or method over another. For example, Brown (1995) made a convincing argument for the case that training with minimal pairs is far less useful than training in suprasegmentals, but he had no data to support the claim. De Bot (1983) also championed this particular view, but each had little or no supporting empirical evidence.

A few attempts have been made to teach pronunciation through suprasegmentals rather than minimal pairs. This usually consists of teaching intonation in a new and original way by displaying the intonation patterns on a screen. The students hear a model sentence and, at the same time, they watch the intonation pattern form on the screen where it is “frozen”. They then attempt to match the model intonation by speaking. As their intonation pattern is also “frozen”, they can compare their production with the model. In this way, students can rely on both auditory and visual stimuli for assistance in the comparison process.

Some studies have looked at the short-term effectiveness of limited training on various aspects of oral production. De Bot (1983), for example, found visual feedback, such as the one described above, effective as a means of improving intonation in Dutch EFL learners. However, he provided only limited training (45 or 90 minutes), and assessed only speech produced immediately after training. De Bot and Mailfert (1982), Gilbert (1980), and Neufeld and Schneidermann (1986) all reported studies with similar limitations.
In contrast with the studies cited above, Derwing and Munro (1997) showed that long-term ESL learners’ pronunciation improved significantly in a 12-week program emphasizing global production skills. In this study, instructors were restricted to use global strategies such as prosodic features, and general speaking habits advocated by Firth (1992) and Gilbert and Rogerson (1993) in their classes. Although the outcome of this study were promising, but because most of the materials still were primarily segment-based, it is hard to know whether concentrating on global strategies will lead to overall improvement in ESL pronunciation.

Macdonald, Yule, and Powers (1994) compared the productions of ESL students under 3 types of instruction-teacher correction, self-study in a language lab, and interactive modification-with those of an ESL control group who received no treatment. They assessed the students’ productions before the experimental treatment, immediately after, and again 2 days later. The results were most positive in the self-study condition. However, there was no indication whether improvement could be sustained over time. Apart from the obvious question over the transfer effect in this study, an interesting fact emerged: that is learners benefited most if they were motivated enough to extend their own learning from the classroom to outside the classroom.

To extend the research cited above, (Derwing et al (1997, 1998) undertook a study comparing the implementation of three conceptions of pronunciation teaching over a longer period of time: one taught with a segmental focus, a second taught with a global focus, and a third that received no specific pronunciation instruction. In this study, the global group received approximately 20 minutes per day of instruction in which the teacher focused on features such as speaking rate, intonation, rhythm, projection, word stress, and sentence stress. The instructor used commercial materials in novel ways. For instance, she would have the students count the number of
syllables and the number of stresses in each line of the teaching materials. The students would tap out the beats and use nonsense syllables to focus on rhythm (as suggested by Celce-Murcia (1996)). The teacher made no attempt to focus on individual consonant and vowel sounds. The segmental group used language-lab materials in conjunction with teacher-led exercises, again for approximately 20 minutes per day. These exercises were designed to improve their productions of individual sounds (Corbett, 1992). Activities included identification and discrimination tasks as well as repetition tasks featuring individual sound contrasts.

Students’ performances in ESL were collected “before and after” the training. They were asked to read sentence-length utterances as well as to produce extemporaneous narratives to approximate a more “real-life” speaking condition (see Derwing 1997; Munro 1995a). English NSs (NS) listeners evaluated the speech samples by rating them for accentedness, comprehensibility and fluency in blind tasks.

The dimensions of accentedness, comprehensibility and intelligibility have already been previously defined by researches (Derwing et al., 1997; Munro & Derwing, 1994, 1995a, 1995b). Accentedness refers to the extent to which a listener judges L2 speech that differs from NS norm. Research has repeatedly shown that even heavily accented speech can be highly comprehensible. Comprehensibility is a subjective assessment of ease or difficulty of comprehension as opposed to a measure of actual intelligibility. The latter refers to how much of an utterance the listener processes successfully; it may be quantified through comprehension questions or an orthographic transcription tasks. Finally, one can also assess fluency through scalar ratings. In this study, fluency refers to overall tempo and flow as opposed to a measure of general proficiency (Lennon, 1990).
The results of the study indicated that for sentence reading task, the segmental group showed the greatest overall improvement in accent after the training process. The training process lasted for 11 weeks. However, after analysing the extemporaneous narratives produced by the three groups, it was found that there was no evidence that the training had any effect on the accent scores. Only the global group exhibited significant improvement in comprehensibility and fluency after the training. At first glance, the results on the improvement in accent demonstrated by the segmental group might suggest that a segmental focus would be marginally preferable to a global focus. However, the results of the extemporaneous narratives showed clear evidence of improvement in comprehensibility and fluency in the global group only.

Derwing et al (1998) argued that pronunciation instruction with only segmental content seemed to require the student to direct attention to the forms in question. When the nature of the linguistic task necessitated attention be divided amongst lexical access, syntactic well-formedness, phonological accuracy, discourse organization and so forth, speakers could not allocate enough resources to phonological concerns for there to be a noticeable transfer of segment-based skills. On the other hand, speakers who had had instruction emphasizing prosodic features such as rhythm, intonation, and stress could apparently transfer their learning to spontaneous production.

This study conducted by Derwing et al (1998) is informative in that it provided some proof that the teaching of suprasegmentals benefits the learners in the long run. It also points to the importance of skill training. In Derwing’s study (1998), the global group teacher did not focus on specific consonant or vowel contrasts. Instead she had the students involved in counting the number of syllables and the number of stresses in each line of the teaching materials and then
tapping the beats and the rhythm of the language involved. To lighten the load of consonants and vowels on their perceptual mechanisms, she made use of nonsense syllables to focus on rhythm (as suggested by Celce-Murcia, 1996). Rather than attempting the impossible feat of trying to cover every stress or intonation pattern in English, she taught them a set of skills such as paying attention to the rhythm of the language which they could use in real face-to-face communication.

The teacher for the segmental group, on the other hand, focused only on the identification and discrimination tasks involving particular contrasting pairs of consonants and vowels. Due to the large number of such minimal pairs, it is hard to see how efficiently learners can use these pairs to help them in their spontaneous conversations. It is conceivable that when the communicative load is high, access skills gained during global language training rather than identifying the errors in their own speech would be easier. Applied linguistics literature still does not provide clear evidence how minimal pair practice can help L2 learners in speech production.

In a sense, as speakers of English have no previous experience with tones in tonal languages, every possible tone pairs as well as some contrasting pairs in Mandarin phonemes would constitute minimal pairs. So the question is firstly: Is it ever possible to cover all the minimal pairs in tone combinations and phonemes between English and Mandarin? Secondly and most importantly, is it ever advisable to teach all of them and will such teaching produce the desired results? So far, no convincing evidence can be gathered from Chinese linguistics research to testify the validity of such training based on contrasting pairs of tones in Mandarin.

Another interesting observation of the teaching behaviour of the global group teacher in Derwing’s study is the way the students were actively involved in the training (Derwing et al.,
They tapped using their hands and so on. This represents a striking contrast to the sedentary nature of the listening tasks the students from the segment group were involved in. It could be argued that it is this active participation on the part of the students that contributed to the long lasting effect of the training on the students’ production in Derwing’s study (1998). A key factor in the success of the training might be that two modalities of learning had been provided by the global group teacher. It can also be argued that the effect of the training might not depend on the number of prosodic patterns or minimal pairs the students had been exposed to. Rather it would probably depend on how many modalities did the students activate during the training process in order to better perceive the language. In the case of the global group in Derwing’s study (1998), the students were involved in a multi-modal experience which involved the audio, visual (tapping), movement, and listening modalities. This experience is likely to be more long lasting and profound in terms of conditioning the body to the English language than just letting learners passively listen to minimal pair contrasts.

3.9.1 Learning through multi-modalities

There is much research that involved the use of multi-modalities in language learning. De Bot (1983) investigated visual pitch feedback with Dutch university students learning English. In a pretest-posttest design, auditory-visual (AV) versus auditory-only (A-only) feedback, and the amount of practice (one session of 45 minutes over 2 sessions) were investigated. After a sentence was presented auditorily, the pitch contour was displayed and learners imitated the sentence. Their pitch contours then appeared on the display below the target. They were allowed to repeat the process a non-specified number of times. The rating of these productions was done by three teachers of English using a 5-point scale. Results indicated that AV feedback
was significantly better than A-only, but the amount of practice time was not significant. The fact that the training under both conditions is short (one session of 45 minutes over 2 sessions) meant that the results were less generalizable to normal classroom conditions.

The benefits of AV versus A-only training (that is two channels of input versus one) have also been demonstrated in improving the perceptual accuracy of non-native sounds for learners of English as a L2 (ESL) with generalization to novel stimuli and transfer to production improvement (Hardison, 2004). The visual stimuli in this instance were full-sized images of speakers’ faces on videotape. However, questions have arisen regarding the effects of brief training sessions, limitation of earlier speech technology systems, and the learners’ ability to interpret displays. These are all the factors influencing a decision to incorporate such systems into pronunciation teaching.

Hardison (2004) carried out a study to determine whether training with computer assisted visual feedback can produce a significant improvement in both levels of the spoken language with evidence of generalization to novel stimuli. This is one of the aims of language training. The study was divided into Experiment 1 and 2. Experiment 1 focused on both quantitative and qualitative aspects of the acquisition of French tones by NSs of American English (AE) using computer-assisted training that permits visual display of pitch contours in real time. The production of the L2 learners was done individually by NSs of French. Each learner’s pre-test, post-test, and generalization of general sentences were randomized. Raters were not told which productions had preceded or followed training.

For the control group, there was no significant improvement. The mean prosody rating for that group declined from pre-test to post-test; and the mean rating for segmental accuracy only
improved slightly in the post-test. For the experimental group using visual feedback, there was an indication that an improvement in both prosody and segmental accuracy was achieved as a result of the prosody training. To determine the extent to which the training would generalize improvement in segmental production and to both prosodic and segmental features of novel sentences, the data indicated that there was an improvement in generalization ability in both prosody and segmental accuracy to novel sentences.

Learners were also asked to fill in an anonymous questionnaire about the training process. The questionnaire responses indicated that the learners had discovered other aspects of French that are equally important in communication. Frequent comments on the rhythm, the use of their voice range, using singing to practice the phrases and so on indicated a heightened awareness of the importance of rhythm in speaking French. These comments are a far cry from the traditional description that to speak a L2 well, one must learn to pronounce the sound of the words correctly.

In Hardison’s (2004) study, Experiment 2 used a memory recall task to explore the relationship between the prosodic information and lexical content of the traces of the training stimuli from Experiment 1 stored in long-term memory by using prosody as the lexical access cue through the use of filtered speech.

In multiple-trace theory, all attended perceptual details of an event are stored in a trace. In other words, prosodic patterns as well as lexical information may be stored together in a trace, and a composite of such traces forms the basis of episodic memory (Goldinger, 1997).
Goldinger (1997) conducted research in the area of memory and retention based on the trace theory originated by Hintzman (Hintzman, 1986). Traces stored in long-term memory are said to return an echo to primary memory in response to a retrieval cue or probe. The content of the echo is the summed contributions of all the traces reacting in concert according to similarity between their properties and the probe. Differing levels of attention paid to the various aspects of the speech signal result in a hierarchical structure of information in the trace such that the strongest element is the one that received the most attention. This hierarchy, in turn, may be influenced by factors such as the perceived relevance of the content to the performance of a task or the uniqueness of the content. The echo can also enhance a probe’s representation by filling in the missing details. Echoes for higher-frequency words do not reflect the detailed characteristics of any particular trace, but they do produce a more generic echo as a result of multiple-trace activation; in contrast, the echoes for lower-frequency words more closely resemble the stimulus, as they are the result of activation of specific old traces (Goldinger 1997). Based on Goldinger’s findings, Hardison (2004) hypothesized that the lexical information corresponding to the prosodic patterns that drew the most learner attention would be the easiest to recall.

Results of experiment 2 revealed that the training group from Experiment 1 was able to recall the exact lexical content of an average of 80% of the filtered sentences. Performance was quite consistent across participants. The best recall results were obtained for training sentences that attracted the attention of learners because of visually and/or auditorily salient prosodic contours (e.g. those showing a substantial pitch range and/or steep slope), expressive lexical content, and features they had identified as particularly difficult for them to produce. Participants from the
control group of Experiment 1 who had not been exposed to the training sentences at all were unable to identify any words.

The research described above agrees that L2 performances are negatively affected by the passage of time. In short, even if students manage fairly well while the model is held in echoic memory, there is substantial difficulty in maintaining that quality of performance. It is as if the model passes poorly into long term memory. This problem is addressed in SEA where the body is utilized not merely as a means by which to enhance perception and immediate performance but as a dynamic mnemonic system, able to sustain the quality of performance over long periods of time.

As shall be explained, computers can also assist in providing mechanisms of redundancy by which to assist in the transfer of material from echoic or short-term memory into long-term memory and so turn a momentary level of expertise into knowledge.

3.10 Computer-Assisted Pronunciation Training

The faster, more powerful computers developed over the last half-decade and the data storage capacity of the CD-ROM has further expanded the educational possibilities of Computer Assisted Learning (CAL). As mentioned before, fewer pronunciation training pedagogies designed for the classroom for teaching a L2 have been successful due to some practical constraints. For instance, from the learning point of view, pronunciation training ideally requires prolonged supervised practice and interaction with NSs with the provision of appropriate feedback at particular junctures on an individual basis. These tasks are extremely time-consuming and difficult to implement in class-based settings. Computer Assisted Pronunciation Training (CAPT) can offer a solution to the problem of practical constraints.
CAPT systems range from CD-ROMs with linked sound files to sentences, systems with a built-in speech analyser and systems with integrated Automatic Speech Recognition (ASR) technology. The provision of CD-ROMs (both data and audio CDs) allows students to access virtually unlimited and realistic L2 input through different channels so that learners can practice individually as often as they wish with unlimited patience from the tutor (in this case the computer program). The integration of either a speech analyser or ASR, however, requires some understanding of speech technology on the part of the teacher.

Just like traditional pronunciation training, these learning environments with built-in CAPT are only effective if they have been designed according to sound pedagogical guidelines. At the moment few of the available systems meet such requirements (Pennington, 1999). One of the most valuable roles for CAPT systems with built-in ASR technology built in is to provide accurate and automatic feedback. Getting good quality feedback is an important aspect of language learning (and learning in general).

Feedback takes place in a language teaching classroom on a daily basis. In a communicative classroom, students are frequently called upon or volunteer to try out a new sentence, conversation or structures. He/she then receives feedback from the teacher or his/her fellow classmates almost instantly in many ways. In the classroom, many channels are used for communicating feedback; for example, through error correction and other channels such as body language, non-verbal behaviour, facial expressions, gestures, tone of voice and so on. Such feedback is usually instantaneous, involuntary (from the feedback provider) and disappears very quickly from the memory of everyone involved. So in a traditional classroom, while we receive a huge amount of feedback on our production, the feedback received seldom
becomes guidance for long term learning in a real, face-to-face communicative interaction outside the classroom. This is due to two reasons. Firstly, this kind of the feedback offered in the classroom clearly has minimal effect on student learning. Secondly, understanding what the feedback contains and how to act on it are not as easy as it seems.

The feedback offered in a L2 language classroom does not contain only information on the correct way of pronouncing or writing something. In some cases, error correction offered can be as detrimental as not offering any feedback at all. An excellent example is in the teaching of Mandarin tones. In a tonal language such as Mandarin, each character has a lexical tone that is quite stable when it is isolated from other words in Mandarin. However, in a sentence situation, a lexical tone of a character is influenced by other characters and their tones before and after it. In other words, a character might lose its stable lexical tone in a larger stretch of discourse. In the Mandarin language, changes in tones across sentences and longer stretch of discourse are very hard to predict and describe. Yet, in every Mandarin as a FL classroom in the world, teachers are constantly pulling students up on their tones by demonstrating the right tones for individual characters and telling them the tone for a particular character is a fourth tone and not the first tone and so on. The problem with such corrective effort is that it is usually ineffective as the immediate context (influence of the words on the left and right hand side of the character in question) of the correction for that particular character is ignored and the effect is usually short-lived and transitory. Furthermore, frequent correction of tones by teachers, can sometime erroneously give learners the impression that somehow they are tone deaf and consequently cannot possibly learn Mandarin.
The advent of computer technology in language learning has added a very interesting dimension to the role of providing feedback. Computer technology can provide an environment in which certain memory traces that work for a particular learner can stay longer within students’ consciousness or sometimes unconsciousness. In a real-life situation, such memory traces can be called up to help facilitate the communication involved. The advantages of feedback offered by a computer are that the feedback is constant; can be repeated over and over again; and allows students to control their own learning.

The design of CALL tools for providing feedback has to depend on the theoretical framework that conceptualizes the tools. For example, in a Speech-Recognition-Based pronunciation training experiment conducted by Tomokiyo et al (Tomokiyo, Le Wang, & Eskenazi, 2000), the feedback offered still follows the model of offering explanation using diagrams of articulatory position and minimal pair practice. Thus the criticisms that are put forward against the various approaches of teaching pronunciation in language teaching are equally pertinent here as well.

Many of the new technologies for CAPT such as the Tell-me-more Chinese (Auralog, 2004), a comprehensive CALL system that provides an overall Mandarin learning system utilizes ASR system for feedback provision. However, when this system was evaluated, many problems were reported. The most pressing concern has to do with the software’s inability to accurately evaluate learners’ production of speech by providing an accurate scoring system, error detection, diagnosis and finally feedback presentation. Ideally, feedback provided should be positively motivating, accurate, easily interpreted by learners and capable of being easily incorporated by learners.
Unfortunately, the ASR system incorporated in the *Tell-Me-More Chinese* CAPT system fails on these accounts. Commenting on the previous version of *Tell-Me-More Chinese*, Zheng (2002) claims that it is very difficult for the students to modify their pronunciation so as to match the model waveform. This is not surprising, while the simultaneous display of the two waveforms in this system may very well be taken as an invitation to produce utterances whose waveform closely corresponds to that of the models. This is, however, not the real purpose of pronunciation training. Indeed, two utterances with the same content may both be very well pronounced and still have waveforms that are very different from each other. Many researchers have expressed doubts on the pedagogical value of these types of displays for this reason. Besides, even a trained phonetician would find it difficult to extract information to correct one’s pronunciation from these displays.

Speech technologists are still trying to find the best measures with which to provide a meaningful score: these should be based on specific pronunciation aspects on which the student can work (but which are difficult to capture automatically) and, at the same time, they should result in a score that is similar to that provided by human listeners. This is where speech scientists come in. Temporal measures, for instance, are strongly correlated with human ratings of pronunciation and fluency. This means that they are able to provide reliable scores for both native and non-native pronunciation assessment but not necessarily for pronunciation training. In a tonal language such as Mandarin, good pronunciation, apart from temporal measures, depend on strong correlates of loudness and intensity with tones. So to create a good and accurate scoring system for Mandarin, the system must base its scoring system on more than just the temporal measure. In Mandarin, error detection is rather problematic. In alphabetic languages such as English segmental features are more important than suprasegmental features.
Segmental errors can be detected with reasonable accuracy by using the right combination of scores. In contrast, an error detection system for Mandarin must be able to detect suprasegmental errors as these errors are the first and primary indicator of a learner’s non-nativeness in Mandarin. However, studies in the field of speech technology indicate that the detection of tonal or intonation errors in any language is underdeveloped at the moment.

Another condition that should be met in order to provide meaningful, human-like or better feedback concerns error diagnosis: ideally, a system should be able to provide a detailed diagnosis of a pronunciation problem and suggest the appropriate remedial steps just like a human tutor. However, recent research on ASR-based CAPT has nevertheless shown that this technology is not yet mature to provide reliable detailed diagnosis of pronunciation errors.

One possible direction to move into in feedback provision is to incorporate recasts, i.e. the correct repetition of a mispronounced utterance without any further explanation, as in most teacher-to-student interactions. This might be sufficient to correct deviant pronunciation in the short term. Another measure that might be incorporated is using the pronunciation clinic idea in which students are given the opportunity to monitor their own articulation (Carey, 2002). This can be done by the computer modelling a target sound or tone within a carrier word, then the student is asked to produce it two or three times, depending on how many attempts it takes to produce an improved production of the sound. Then the student is asked which sound they think is better, the first, the second, or the third? Research shows that by monitoring their own articulation, the student can often perceive or feel which utterance was closer to the model (Carey, 2002). By asking the student to reflect on what they feel is a fair production of the
model, more useful feedback can be obtained than if the computer just comments “good or better or try again”.

Finally, providing meaningful feedback means providing feedback that can be interpreted by the learner. In other words, explanation is a function of meaning and meaning is personal. This means that all the information obtained in the speech recognition phase of the process needs to be combined with information from other possible sources in order to present to the student in a clear and unambiguous way. At the moment, ASR-based CAPT systems such as the one presented in *Tell Me More Chinese* (Auralog, 2004) do not tell the learner what they should do to interpret the feedback. Therefore, after repeating the highlighted portion of the sentence many times and perhaps with each attempt getting more and more correct, learners cannot get a concrete idea on how to transfer the learning received during these attempts with this particular sentence, to future attempts with other sentences outside the domain of the 200 sentences prescribed by the system. Learners, for example, will not know how to manipulate their body posture, movements or even their articulatory systems to ensure that their future attempts are successful. In a sense, when the learner repeats the same sentence next time, again it will be a trial and error process as they cannot recall what they learned last time they used the system. In other words, the learners are doing the exercises but they are not learning.

At the moment of writing this thesis, a pedagogically sound software program for teaching the pronunciation of Mandarin like *Streaming Speech* (Cauldwell, 2004) for English is not available. The stated objective of *Streaming Speech* is to use “authentic fast spontaneous speech of NSs of English…to teach listening and pronunciation in a revolutionary way” (cited in Lian 2004). The theoretical framework for this resource is based on the assumption of perception
comes before production. While not stated explicitly, it is also implied that perception and production are viewed as mutually reinforcing processes and the workings with the materials in the resource results in a cyclical process of progressive refinement of both. This orientation is precisely what this thesis adopted. The innovative aspect of this resource is the focus on the prosodic features of a speech unit in English. For example, a special notation is used to identify the stressed syllables and which is arranged in the shape of an intonation curve. Feedback is also provided to the students if they click “highlight a portion” on the speaker icon to get the pronunciation of a highlighted syllable. Students can also record the final product of the “fast speech” in either the “fast speech” format or in the “dictionary” format. As Lian (2004) pointed out this “gives them opportunities for experiencing contrasts not as listeners but as producers. These changes in forms of perception are designed to give learners the opportunity to listen and perceive differently, thus increasing the probability of changing the ways they both hear and produce.” Again, the danger of letting students detect gross errors in their performances surfaces again. However, it is extremely good to see that the design of pronunciation teaching software has finally recognised the importance of prosodic features of a language even if to start with in English only.

Though not a fully-fledged commercial product like Streaming Speech for English, the course materials used in the study also included a speech analysis system- the Sptool (Zhang & Newman, 2003). It was designed to give learners the opportunity to listen and perceive differently, thus increasing the probability of changing the ways they both hear and produce. In the present study instead of adopting an ASR-based CAPT system to enhance learning and expect such a system to provide accurate and informative feedback to beginning learners of Mandarin, a speech analysis tool (Sptool) for tone languages was developed and tested with the
experimental group of students. The decision to create a speech tool specifically for tonal languages was made because the existing speech tools such as Winpitch (Martin, 2003) and Praat (Boersma & Weenink, 2003) are difficult for learners to use as they were initially created by phoneticians for phonetic analysis. By contrast to other ASR-based CAPT system, the speech analysis tool Sptool (Zhang & Newman, 2003) is designed to offer feedback that is non-judgmental, i.e. without a scoring system, and allows students to explore and reflect during the process of learning, and not just at the end of the learning process. When using the tool, students can listen to the language and reflect on it by observing visually the differences between their production and the NS model. In addition to this resource, students are also taught physical ways of enhancing memory traces of the language learned in the classroom context. In this way, students can act upon their perception and change the process of production in their private study. A detailed description of the Sptool and the classroom teaching procedure will be provided in chapter 4.

3.11 Essential features of a Somatically Enhanced Approach in the creation of a Mandarin language learning environment

Several pedagogical implications for the teaching of Mandarin tones can be drawn from researches on first and SLA that have been reviewed in this chapter:

1. More emphasis should be placed on pronunciation by teachers.

Most Chinese textbooks for adult learners devote little attention to phonetic explanations and provide few phonetic exercises, usually confining these to the preface or the beginning lessons. The minimal phonetic instruction available in standard textbooks is usually focused on the articulation of segmentals, presented in the form of repetition drills and minimal pair drills. As
reviewed in Section 3. 4.2.1, the use of tonal contrast drills in the language laboratory is theoretically flawed and can be counter-productive for beginning L2 learners. One belief in the field of English language teaching is that pronunciation should be integrated into the general language lesson instead of being taught as a separate subject (Brazil, 1994). This is a belief that is shared by the present researcher and implemented in the present study.

The studies summarized in Section 3.3.5 indicate that teachers’ attitudes can be crucial to the success of the tone acquisition of their students. In particular, what teaching materials selected by the teacher and how she/he corrects the L2 learners are important. The review in this chapter has offered us proof that presenting teaching materials in their realistic authentic contexts with the intonation, stress, and other prosodic patterns intact may help L2 learners in a multitude of ways. In fact, presenting new language this way, not only would facilitate the acquisition of the language sounds, it would also facilitate the exploration of the syntactic system of the new language as well.

2. Need to increase the exposure to Mandarin in a learning of Mandarin as a FL environment.

The literature review in this chapter also shows that massive and continued access and exposure to Mandarin input is instrumental in improving L2 learners’ perception and production of Mandarin tones. Limited results from studies conducted on the acquisition of tones in learning Mandarin as a L1 show that no reliable generalizations can be made regarding a natural order of tone acquisition. In practice, the order of acquisition seems to be very much determined by what the child is exposed to. Consequently an optimal learning environment would be one where exposure to Mandarin can be accessed easily and in large volume.
Access to a rich Mandarin language input environment must also be accompanied by ways of increasing learner motivation so that a learner is pushed forward greatly in the acquisition of a L2 (Klein 1995). This can be done by lowering the affective filter as well as allowing learners to gain useful feedback and gain a sense of achievement in the process of learning.

3. Emphasizing the exploration of learners’ vocal range and the body in the process of language learning.

Many students do not know where to start and end a tone. This phenomenon is noticed also by Miracle (1989) and Shen (1989b) who argued that the major cause of errors for L2 learners of Mandarin is tone register rather than tone shape. Familiarization with tone register means knowing one’s own voice range well. Shen’s recommendation of not using tone diacritics in teaching illustrates the need for learners to concentrate on the exploration of their own voice range. Zhao’s vocal-cord method provides us with an idea of what kind of exercises might be useful in enhancing learners’ experience of their own pitch range with regards to the four tones (Zhao, 1988).

4. Provision of Mandarin language input through a multi-sensory environment

In section 3.2.1.5, Lin (1985) described a series of extremely useful ways of providing input. Most of the suggestions will be taken up in the present thesis except the use of musical scales. This is because it is unrealistic to expect classroom teachers to produce all their teaching materials in this form week by week. Not only such efforts might be time wasting, they might also be ineffective as well. However, the provision of a visual component for the teaching
materials is acknowledged as important. In this thesis, this aspect of input provision is accomplished through the use of a speech process tool built into the teaching materials.

The research findings obtained from learner-centred approaches described in sections 3.9 also demonstrate that:

1. Focusing on prosodic features such as rhythm, intonation and stress can enable the transfer of learning gained in the classroom to spontaneous speech production (Derwing 1998).

2. Learning through spoken, aural, and modality involving movement is more efficient than doing passive listening exercises based on minimal pairs.

3. Training with computer-assisted visual feedback can produce a significant improvement in both levels of the spoken language with evidence of generalization to novel stimuli (Hardison 2004).

4. Visually and/or auditorily salient prosodic contours with expressive lexical content make the recall of the training sentences easier.

5. The Importance of output

As reviewed in Section 3.3.3, adult learners of Mandarin as a FL and children learning Mandarin as a L1 differ in the amount of cognitive strategies they can use in the process of learning. While there is little communicative demand on children learning Mandarin as a L1, the communicative demand on adults is enormous thus their need to produce output in the TL as soon as possible. Furthermore, as Swain (1995) argued, producing output in the TL is extremely
beneficial for improving learners’ fluency and accuracy in the TL and is essential in allowing learners to interrogate their IL development. Therefore, speaking practice and performance are essential in the FL environment.

6. Catering for the predictable and unpredictable needs of learners

The review of the socio-psychological factors in language learning reminds us that the classroom is an extremely complex site of learning with different learners’ individual needs constantly desired to be fulfilled. Therefore, instead of adopting the losing strategy of attempting to cater for everyone’s predictable and unpredictable needs (which is impossible in reality), the environment under investigation in this thesis attempts to create a learning context in which learner’s level of inhibition is reduced.

3.12 Summary

The literature review conducted in this chapter contains a great deal of information from a number of fields of study on a number of mechanisms or procedures that might benefit L2 learners in their attempt to learn a FL. However, evidence testifying that one particular way is superior to any other is generally absent. This thesis contends, however, that it is not necessary, on a concrete level, to search for a set or configuration of mechanisms or procedures that worked for everybody in order to enable meaningful learning to take place. To benefit learning, it might be more fruitful to adopt an abstract principle which allows the learning to take place that is determined fundamentally by the learner but helped by the teacher. This means that the environment needs to enable individualized learning to take place using a variety of tools and mechanisms for learning.
The suggested language-learning environment focuses on the teaching of Mandarin tones and prosody involving a SEA and uses computer-assisted visual feedback. This thesis contends that such an environment is more likely to achieve better results in learners’ L2 speech production than the traditional training with minimal pairs. The aims of this thesis are first of all to construct such a learning environment and then to investigate the efficacy of such an environment. In the next chapter, a detailed description of the study is presented including a detailed description of the SEA involved.
Chapter 4 Toward a Somatically Enhanced language learning environment for teaching Mandarin tones and prosody

4.1 Introduction

The characteristics of the Somatically Enhanced Approach (SEA) for teaching Mandarin tones and prosody proposed in this study are as follows:

1. An important aim of the course in beginning Chinese is to weaken the phonological sieve of one’s native language, so comparison of new sounds in Mandarin with sounds in English is avoided at all cost. Therefore, the language material used is not accompanied by any kind of romanization. The material is presented in Chinese characters.

One of the major barriers of adult learners learning Mandarin as an L2 is most likely to be L1 interference particularly where the learner is a NS of a non-tonal language such as English. An alphabetic language such as English uses rules that do not coincide with Pinyin. In using Pinyin, NSs of languages such as English or French, are influenced, not merely by interference of their L1 but by the design of Pinyin that, while it uses the same letters, does not put them to the same purpose. In short, if L1 interference were considered the primary barrier to communicative competence in learning Mandarin, then the representation of Mandarin by the Roman alphabet can only be construed as an enhancement of this difficulty. Traditionally the teaching of Pinyin has been imagined as a bridge but it may in fact be a wall.

Traditionally, Chinese characters are presented usually with Pinyin as well as tone diacritics above the Pinyin. Learning through such representations, students are encouraged to read Mandarin utterances in Pinyin with the tone diacritics to remind them of the tones. Students often if not invariably make mistakes, not merely with
respect to tones, but even with consonants.

2. “Helpful” tools to remind students of tones such as tone diacritics are not provided in SEA. This is because the design of the course is to encourage a direct experience of the word of the TL as suggested by Steiner (Steiner, 1996). Diacritics above the text are designed to indicate which of the four tones of Mandarin is to be enacted in the performance of that syllable. However, the four tones in Mandarin are more of a pedagogic model of the phonology of Mandarin than of nature. While such marks may be helpful they do not actually represent the fluidity of Mandarin tones. For example, in the case of Tone 3 with its complicated tone sandhi rules, tone diacritics would probably misguide students as often as they assist them especially at the very beginning of learning Mandarin (Jiang, 1996). Students learning Mandarin through diacritics would probably strive to reproduce the prescription of these marks regardless of their actual experience of the language. Past research showed that remembering diacritics or reading diacritics would not necessarily help in the production of the TL.

3. Intellectualising about Mandarin (in terms of explaining grammar and translation) is kept to a minimum in the two-hour sensitization session. As grammar translation is likely to be helpful in raising students’ awareness about many aspects of the Chinese language and culture, discussions about grammar are carried out during the subsequent tutorials. The lecture notes contain explanation of grammar and other information about the Mandarin language which students can read at their leisure. However, in keeping with the spirit of SEA, the focus of the grammar explanation is not on what the grammatical patterns are but on when and how a particular sentence pattern is used and
how the meaning of that sentence is particularly influenced by the particular grammar pattern. In other words, the explanation and illustration of grammatical patterns are always conducted in their relevant contexts. Translation is kept to a minimum so as to avoid over analysing the L2 in the learning process.

4. The language is studied globally in that the isolation of single sounds or the systematic cutting up of syllables is avoided when modelling the language to students. Prosodic features such as intonation patterns, stress, and intensity are taught and respected in relevant contexts.

While it is indeed true that a particular syllable is said with one of the four tones, a tone is profoundly modified as one tone flows onto another and is inflected by grammatical or personal mood, e.g. imperative, interrogative and so on. Therefore, while being able to pronounce lexical tones correctly might be acceptable in an academic context, the ability to produce language with correct tones and rhythm and a whole host of other prosodic features in Mandarin (some of them are still unknown or may be unknowable) generates status among NSs who are more impressed by their perceived naturalness of the learners’ performance. In other words, what SEA attempts to do is to develop a “feel” for the Mandarin language which enables them to produce spoken language which is acceptable to NSs of Mandarin.

5. Another characteristic of SEA is the use of CD-ROM technology in the provision of learning materials.

The Data CD and audio CD have replaced the traditional tapes because they are flexible to use. They can contain a number of multimedia objects such as sound files, texts, video
segments and accompanying worksheets. The audio CD provides easy access to all the sound files associated with subject materials at a click of a button. Such easy access and portability invariably increases students’ learning time outside class. Access to the sounds of vocabulary, sentences and dialogues are provided on both the course data CD and audio CD so that students can listen to the relevant sound files in cars or on Ipods and so on. The teaching material for this subject has been divided into twelve units. Each unit covers some of the language used on a particular topic. For example, the first unit covers the topic of “Introduction” in Mandarin (see Appendix I). Within each unit, all the vocabulary, sentences and dialogues on the data CD have been attached to their own audio files which can be listened to on their own or through the speech processing tool (Sptool) and a filtered audio file. Students are also provided with the printed version of the text file for their own note taking through the University bookshop. Students are also provided with an audio CD with all the associated sound files transformed into tracks with corresponding track names. The audio CD enables students to access the learning materials through an ordinary CD player or Ipods.

The provision of data CD and audio CD is designed to make the preparation and review of the specific learning content each week in their individual study outside class easier. Students are also free to create any alternative pathways as they deem fit. The teaching materials are designed around a number of activities which are designed to train students’ perception to the phonology of Mandarin.

6. Students are also provided with a speech processing tool (Sptool) so that instant feedback can be obtained by comparing their own utterances with that of the NS model.
4.2 Phonological Model

In his seminal work *The Principles of Phonology*, Trubetzkooy defined the tasks of phonology as how the infinite variety of physical sounds fit in with the finite nature of language structures that govern a particular language (Trubetzkoy, 1939). In language acquisition research, the term “structure” has been used in a variety of ways. As a noun, it is usually used in conjunction with “language” or “grammatical” to show that the aim of learning, through the result of an internal activity carried out by the learner upon meeting language stimuli, is to reduce the infinite amount of random language stimuli into some kind of finite, manageable structures within a particular language. As a verb, it is used to describe the way language behaviour is a result of organisation through the structuring activity both at the level of production and at the level of perception.

As this thesis is centrally about how L2 learners learn to organise L2 language input at the beginning of the L2 learning process, “structure” is mainly used as a verb to describe what and how human beings deal with new L2 language input. For instance, in production, the learner must structure the non-linguistic experiences that he/she wants to talk about so that the available extra-linguistic and linguistic means can be applied to it. As far as perception is concerned, structuring happens when a learner is called upon to select different acoustic stimuli through the senses in order to integrate the message. In other words, in terms of perception, the structuring activity implies re-ordering, usually in an unconscious way, by selective filtering of redundant data that are usually perceived globally. This structuring activity is a dynamic process.

Before we describe the structuring activity in detail, it is essential to understand how a sound of speech is made up. For instance, a vowel is physically made up of a complex set of many frequencies produced simultaneously. When a vowel is heard, everything it contains is heard.
However, when listening to a sound, it is not necessary to catch all the elements in order to recognize it as recognition only requires some of the sound spectrum.

If, for example, the sound [i] in French is recorded and listened to successively through octave filters:

```
between 150 and 300 cps, we hear [u]
between 300 and 600 cps, we hear an intermediate sound between [u] and [o]
between 400 and 800 cps, we hear [o]
between 600 and 1,200 cps, we hear [o]
between 800 and 1,600 cps, we hear [a]
between 1,200 and 2,400 cps, we hear [e]
between 1,600 and 3,200 cps, we hear [e]
between 2,400 and 4,800 cps, we hear [i] lax
between 3,200 and 6,400 cps, we hear [i] tense — Cf. R. Renard,
```

“L’appareil Suvaglingua, instrument de recherché et de correction phonétique”, R.P.A., 4, 1967, notes 13 and 14, pp. 62 and 63. Taken from (Renard, 1985)

According to the above table, the French [i] is made up of all those sounds, each sound occupying a different frequency. Yet when French speakers were asked to identify these sounds, all the sounds were identified by French speakers as productions of the phoneme /i/. This manipulation of the French [i] sound showed very clearly the superabundance of the acoustic reality. Through perception a distinction between what is necessary for [i] to be recognised ([i] is recognised between 3200 Hz to 6400 Hz, previously in cps: cycles per second) and what is superfluous in the plethora of information which comes through the vocalic timbre (the quality of the vowel) was made by French NSs.
However, when it comes to a L2 learner learning French, the learner would perceive a sound through hearing all the frequencies that are contained in a sound. As the learner’s perception is likely to be mediated through the pre-established code based on his/her mother tongue, when it comes to perceiving a sound in L2, he/she is likely not to recognise the sound /i/ at the frequency recognised by a French person. He/she is likely to recognise the sound /i/ at a frequency dictated by his/her mother tongue such as between 300-600 Hz. Thus, he/she is in danger of confusing /u/ (between 300-600 Hz) with /i/ (between 3200 and 6400 Hz), or with /o/ (between 400 and 800 Hz) as these sounds also occupies part of the spectrum that contains /i/ at different frequencies.

A remedial strategy could be, in order for L2 learners of French to perceive [i], to eliminate all the sounds at frequencies above 300 Hz and only leave frequencies between 3200 and 6400 Hz through a process of filtering so that L2 learners can be exposed to the French [i] at the correct frequency. This strategy is not very different from the strategy adopted by McCandliss et al in training Japanese speakers to discriminate between the difference between [r] and [l] in English (McCandliss et al., 2002) in which exaggeration was used to highlight the difference.

A filtering process was also used in the Verbo-tonal method (VTM) of phonetic correction, developed at the Institute of Phonetics of the University of Zagreb by the late Professor Petar Guberina. This method of corrective phonetics had been used by teams of teachers and researchers to write structuro-global audio-visual courses under the supervision of Professor Guberina and Paul Rivenc in the 1960s (Heras & Regan, 1985; Renard & Vlasselaer, 1976). Filtering in VTM (Renard, 1975) is a load lightening measure through which only the “relevant frequencies” for a particular sound, in this case, that of /i/ in French, is allowed to remain.
As noted before, certain practices of SEA owe a great debt to VTM. These practices, derived in some part from VTM, provide an alternative approach to education in Mandarin tones and prosody. Instead of writing things down in Pinyin and putting little marks above consonants and vowels, VTM offers to educate students in the performances of phrases in action. VTM does not offer students the complete mastery of the infinite diversity of the NS speech. However, students trained with the benefits of VTM might be better prepared to perform a phrase in such a way as to be readily intelligible to NSs of the TL community.

VTM was initially developed for the rehabilitation of the deaf. The aim of VTM was to teach people how to hear and produce acceptable language in their mother tongue using their residual hearing. Critical to this success is the ability of the deaf subject to reproduce intonation patterns of the TL. Such intonation patterns cannot be deduced from the written representation of the TL. Deaf people must learn to pronounce the phonemes of their TL replete with variations without the aid of some system similar to Pinyin.

Normal, hearing students learning Mandarin as a L2 are not so different from the deaf subjects of Guberina’s studies. Research reviewed in Chapter 3 found that L2 learners of Mandarin had little trouble with consonants and vowels but a great deal of difficulty with tones. In fact, the acquisition of consonants is even easier if Pinyin is not used (Zhang, 1998). VTM never claims to guarantee students a complete mastery of any language because many aspects of language and language use of any language have not been defined and may be not definable. However, through the use of pedagogical method derived from VTM, students may be expected to produce utterances that are more readily intelligible to native speaking audience albeit it is flawed.
4.3 Classroom methodology

The learning of pronunciation involves both perception and production. Though some have advocated the advantages of delaying language production (Gary, 1975), there is substantial amount of evidence from the fields of psychology and psycholinguistics to suggest that the urge to imitate is so overwhelming for social cohesion that human cannot perceive and not act especially when it comes to language learning (Bargh & Dijksterhuis, 2001). In other words, in L1 acquisition, children are perceived to be linguistically non-productive because their vocal apparatus are not mature enough to produce language. However, if production is expanded to include all manners of imitative behaviour, then even the “non-productive” children can be seen to be constantly productive.

In SLA of adult learners, their perceptual and productive systems have obviously been trained to perceive and produce their mother tongue. Therefore, when they start to learn a L2, students come with a kind of “deafness” (due to the L1 phonological sieve as postulated by Trubetzkoy) to the sounds of a FL. The groups of adult learners under study in this thesis are also adult learners who have passed through the sensitive period (SP) for language acquisition (Werker & Tees, 2005). Though it has been widely accepted that the SP is not unmodifiable, it is acknowledged that modification is more difficult after SP (Tees, 2001). That is to say, in the case of teaching adults to speak a tonal language, it is necessary to defeat the learners’ L1 phonological sieve.

As perception and production might be directly related to each other in an input-output loop, students need to practise intensively their newly-learnt articulatory patterns in order to develop a “feel” for them and at the same time fixing in their long-term memories acoustic models of the
sound(s) being learnt. The classroom methodology describes essentially the range of procedures used to highlight the prosodic aspects of Mandarin in the sensitisation session.

4.3.1 Sensitization Session

In a five hour per week course, this session takes place during the first two hours of the five-hour contact. This arrangement is purely administrative. It should normally occur in a room, preferably carpeted, where tables, chairs, etc., can be removed in order to leave as much space as possible. Language laboratories should be avoided at this stage for reasons previously mentioned in Chapter 3, section 3.4.2.2.

4.3.1.1. Step 1: Relaxation

The first step in the learning process is a relaxation procedure adapted from the success of relaxation techniques used in language learning in the 1980s. In the early seventies, Lozanov method of language-learning (Suggestopaedia) first became known in the Western countries. It was originally developed by a Bulgarian physician and psychotherapist, Georgi Lozanov. It has been used successfully in the U.S.A., Canada, France and other European countries. It was claimed that it could speed up learning by some fifty times. In America, however, it was reported that the rate of learning was only 2.5 times better than under ordinary teaching conditions (Bancroft, 1978). Furthermore, as reviewed in Chapter 3, section 3.3.5, there is also biological evidence that suggests that a fun or playful atmosphere is good for learning because it actually influences the architecture of the brain.

As the classroom is also a social site of learning, lowering the learner’s level of inhibition can also make the learners’ egos more permeable (Guiora et al., 1972b). This relaxation step also is
designed to reduce the language shock experienced by many learners especially when they are required to speak in the TL.

The first step in the sensitization session is to ask students to lie on their backs on the floor and if possible, with the classroom darkened, then carry out mind-calming exercises for some five to ten minutes. This allows them to be more relaxed and therefore more receptive to the language input. The following is a description of the relaxation phase of the sensitization session conducted in the study:

It is 5.30 pm on a Tuesday afternoon, in a large room capable of holding up to 50 students; the lecture chairs with attached arms have been pushed to the perimeter of the room. The students are randomly slouched on their chairs randomly, relaxing after a tired day of either work or lectures.

The teacher walks into the room carrying the necessary computer gear, CD-ROMs and so on. She greets the class cheerfully with “nǐmen hǎo” (hello, everyone) and puts the CD-ROM in the computer. “Now, leave your seat and lie comfortably on the floor and listen”. Then the following audio file is played:
Imagine that you are lying on your back on the grass on a warm summer day and that you are watching the clear blue sky without a single cloud in it (pause). You are lying very comfortably, you are very relaxed and happy (pause). You are simply enjoying the experience of watching the clear, beautiful blue sky (pause). As you are lying there, completely relaxed, enjoying yourself (pause), far off on the horizon you notice a tiny white cloud (pause). You are fascinated by the simple beauty of the small white cloud against the clear blue sky (pause). The little white cloud starts to move slowly toward you (pause). You are lying there, completely relaxed, very much at peace with yourself, watching the little white cloud drift slowly toward you (pause). The little white cloud drifts slowly toward you (pause). You are enjoying the beauty of the clear blue sky and the little white cloud (pause). Finally the little white cloud comes to a stop overhead (pause). Completely relaxed, you are enjoying this beautiful scene (pause). You are very relaxed, very much at peace with yourself, and simply enjoying the beauty of the little white cloud in the blue sky (pause). Now become the little white cloud. Project yourself into it (pause). You are the little white cloud, completely diffused, puffy, relaxed, very much at peace with yourself (pause). Now you are completely relaxed, your mind is completely calm (pause), you are pleasantly relaxed, ready to proceed with the lesson (pause) (Bancroft, 1978:174).

This constitutes the relaxation phase of the classroom procedure. As Lian noted (Lian, 1980:16)

Relaxation of the body will bring about a lowering of conscious and unconscious resistance to the learning of a FL. Speech and the production of sounds appear to be the result of the muscular behaviour of the body as a whole which, with appropriate reinforcement, has given rise to a number of set patterns of muscular contractions. If these still operate when one attempts to learn the articulatory patterns of a FL, then the resulting articulatory sequences will be deformed, sometimes beyond recognition.

It therefore becomes very important to reduce the influence of the set of English or individual muscular tensions and movement to a minimum when learning Mandarin. Relaxation techniques appear to be an effective way of reducing, if not eliminating, such conditioning so that it can be replaced with another set of muscular tensions and movements: those of Mandarin. For this reason, the relaxation phase of the course is extremely important. Relaxation is therefore the first step recommended for the teaching of tones.
4.3.1.2 Step 2: Humming

In the studies conducted for the rehabilitation of hard of hearing children and adult, SUVAG machines to transmit language sounds at low frequencies without distortion were used. With the advance of computer technology, the production of language sounds at low frequencies can be done easily through the application of the correct filters. For instance, in Professor Lian’s study on teaching French intonation, filtered sentences were produced by eliminating all sound elements above 300 Hz so that only the intonation of the sentences was left. In this particular study, though the same filtering technique was used to produce the sentences in the teaching materials, no recording of sentences at low frequencies was played in the classroom procedure because good quality amplifiers and loudspeakers were not available for this purpose. Ideally, in order to sensitize student’s perception to the tones of Mandarin, filtering should have been used to highlight the prosodic elements of the language. Due to the lack of such suitable equipment, instead of filtering, humming was used. However, the “humming” procedure used in this study is not the same as the filtering process recommended by VTM (Guberina, 1970, 1972; Renard, 1975) as it still retains all the frequencies but loses the intonation pattern of the original utterance. The intonation pattern that is obtained through humming is a simulation of the original intonation pattern of the sentence. Filtering out the frequencies above 300 Hz as recommended by VTM, on the other hand, eliminates the potentially interfering consonants and vowels and leaves the original intonation intact in all its detail. Nevertheless, by humming, the suprasegmental aspect of the utterances is highlighted at the expense of the vowels and consonants. Therefore, though less ideal than the use of filtered sentences, humming nevertheless allow students to focus on the suprasegmental aspect of the language. The filtered sentences, however, are actually provided on the data CD for the course.
“Now, get up and stand in a circle.” The teacher joins the circle. The teacher says “I will hum to the sentence and please hum with me while walking slowly in a circle”. This is done for 5 times. (Step 2)

Step 2 involves humming to the intonation of the sentences without the vowels and consonants (5 times). This is used to highlight the intonation and tones of Mandarin. Intonation here refers to the sound changes produced by the rise and fall of the voice when speaking.

At this stage students take an explicitly active part in the proceedings through humming along to the model. They listen to the hummed model (of the intonation) by the teacher. Then they are asked to repeat by “humming along” to the intonation. This is a way for them to produce an uncluttered sound string free from interfering vowel and consonant sounds. The removal of vowel sounds is particularly important for English speaking learners of Mandarin because it forces them to prioritise the tones and prosodic aspects of Mandarin. As the input and output of the language uttered mutually reinforce each other, three factors must be considered in order to maximise the benefit of production and perception. First, such a structure should be relatively easy to produce a maximum of 5 to 7 syllables. Second, humming allows them to develop a much better perception of the melodic patterns concerned. Third, this, like every other subsequent phase, would serve to reinforce the preceding one.

The intonation of the sentence to be hummed was repeated 5 times. This phase was still very much centred upon sensitising students through the repetition of a model.
4.3.1.3. Step 3: Clapping to the rhythm of the sentences

“Now, I will clap to the rhythm of the sentence and then you can clap after me while walking in a circle.” (Step 3) This is done for 5 times again. The intonation of the sentence is again hummed in this fashion while the clapping is taking place.

The students, while listening to and “feeling” the intonation patterns, begin to move in harmony with the rhythm and intonation of the sentences modelled by the teacher. The teacher provides the beat and the rhythm of the sentences according to the stress and discourse features of the sentences. For instance, in teaching the sentence “nǐ jiào shénme míngzi? What is your name?” when the same sentence is presented in character form, there is no space separating the characters, so students reading the sentence do not know which character goes with which unless they know the meaning of the characters. For instance, in spoken speech “shénme: what” always go together. If a learner only learns this through reading then it is highly likely that he/she will always introduce a pause in between “shén” and “me” and another pause between “míng” and “zi”. However, in prioritising the spoken over the written language in this course, the teacher demonstrated the beat of this sentence by providing a beat for that group of words in the following manner:

[nǐ] [jiào] [shénme] [míngzi]?

1 beat 1 beat 2 beats 2 beats

The clapping to the intonation patterns created a rhythm that students could follow while walking in a circle. This also allowed students to observe and experience how stress, realized by length and loudness in Mandarin, was tied to meaning. This also allowed them to observe the key words in a sentence and realize that not all words were of equal value and that in making
oneself understood, some words were more important than others. This training was also
essential in equipping them with the strategies of prediction and advanced planning in listening
comprehension.

4.3.1.4. Step 4: Incorporation of movement and gesture

According to Birdwhistell, 65% of communication is non-verbal and only 35% can be
attributed to the verbal element (Birdwhistell, 1970). These findings suggest the great
importance of movement and gesture in communication. From studies carried out, particularly
in the U. S. A., it has been shown that movement is very closely linked with precise events in
rhythm and intonation (Condon & Ogston, 1971). This is known as “self synchrony”. “Self
synchrony” refers to a process whereby the body of a speaker moves closely in time with
speech (Condon & Ogston, 1966). Self synchrony is not simply confined to hand gestures.
Movements of all parts of the body are also closely synchronised with speech, particularly in
terms of vocal stress (Halliday, 1970). For instance, spoken English is produced in groups of
words, typically averaging about five in length, where there is only one primary vocal stress,
conveyed principally through changes in pitch, also through changes in loudness or rhythm
(Bull & Connelly, 1985). In one study of student conversations it was found that over 90 per
cent of primary stresses were accompanied by some kind of closely synchronised body
speech, is never disembodied but is always manifested through behaviour” (Condon & Ogston,
1967).

One of the major problems encountered by FL learners is that they attempt to utter sentences in
the FL while, at the same time, unconsciously preserving the set of movements that normally
functions in their mother tongue (Gassin, 1990). A synchrony of this nature regularly leads to
failure to reproduce the appropriate stress patterns of the FL in question. Relaxation and mind-calming exercises in step 1 are meant to reduce such muscular conditioning. The humming and walking to the rhythm in steps 2 and 3 are also designed to develop synchrony of the body with the TL.

Research has shown that experience learning a tone language is necessary for categorical perception of tone distinctions (Gandour, Wong, & Hutchins, 1998; Van Lancker & Fromkin, 1973; Wang, Jongman, & Sereno, 2001 cited in Werker & Tees, 2005). Moreover, direct experience in learning a tone language results in left-hemisphere lateralization for perception of tone contrasts. Non-native tone contrasts such as intonation patterns fail to reveal specialized left hemisphere processing (Gandour et al., 2000; Gandour et al 1998). Research on American bilingual speakers of Mandarin show that Chinese speakers have a much wider voice range when speaking Mandarin than English speakers speaking English (Chen, 1974). Furthermore, American L2 speakers of Mandarin speak Mandarin with a wider voice range than when they speak their mother tongue. This indicates that the experience of learning a tone language such as Mandarin should actually manifest in numerous physical changes in the lateralization of the brain and in the voice qualities of the speakers. Apart from these physical changes, what mechanisms in the body of a speaker are responsible for accommodating the changes in pitch, loudness and rhythm of a language? One of the mechanisms especially in terms of pitch proposed describes how different muscular effort changes the production of sounds. This relationship is encapsulated by the concept of tension in acoustic phonetics.

4.3.1.4.1 Tension

Tension is a term used in phonetic classification of speech sounds, referring to the overall muscular effort used in producing a sound. The contrasts are labelled variously, e.g. fortis
versus Lenis, tense versus lax. This contrast is viewed as particularly important in distinctive feature theories of phonology, where tense is one of the main features set up to handle variations in the manner of articulation. Tense sounds are sounds produced with a relatively strong muscular effort, involving a greater movement of the (supraglottal) vocal tract away from the position of rest (cf. fortis) and a relatively strong spread of acoustic energy. The opposite term in Jakobson and Halle’s system is lax (Crystal, 1997).

The following factors contribute to conveying the impression of greater tension:

a) in the case of consonants (or, to be more precise, of the consonantal elements of language) a gr
b) in the case of vowels (or the vocalic elements of an utterance), on the one hand, greater tension in certain laryngeal muscles and/or the vocal cords which is combined or not with an increase in air-pressure leading to a relative raising of the pitch of the fundamental, and on the other hand, a deformation of vocal tract and a degree of rigidity in the tongue body, which is relatively greater and of longer duration, thus bring about a change in timbre and a lengthening of the vowel sound (Renard, 1975:62).

Though Verbotonal studies mainly focused on vowels and consonants in various languages, the concept of tension has also been applied to tones. Zhao (1987) described the four tones of Mandarin in terms of varying degree of tenseness of the vocal cords. For instance, in order to produce Tone 1, the vocal cords should be kept tense; to produce Tone 2, the vocal cords at first neither tense nor lax, then tense rapidly; to produce Tone 3, the vocal cords become lax immediately after being tensed, and then tense again; to produce Tone 4, the vocal cords suddenly tense, and then lax gradually.

Thus the corresponding gestures have been developed to produce the various tensions of the four tones are as follows:

Tone 1: requires the vocal cords to be tensed and to be kept tensed.
In order for students to experience the tensing of the body tension when pronouncing the Tone 1, upward movements are very important because they set up tension which allows students to perceive and generate the language through getting the feel in their body for the production of this tone. This can also be achieved by having very tense hands, with the fingers spread out and the palms facing upwards. The elbows should be vertical and held close to the body. Students then push upward as though trying to touch the ceiling and keep to this posture, when pronouncing Tone 1.

Furthermore, as Tone 1 starts at a higher frequency with the tensing of the body greater than what most Australian speakers are used to, extra physical efforts need to be made to remind one that one must start high. To stretch one’s muscular system to express these Mandarin tones, one must not slouch in seats. By asking students to stand up straight and walk in a circle with various gestures, students were experienced the coordination and synchronization of various muscles with the sounds uttered.

Tone 2: the vocal cords are at first neither tense nor lax, then become tense rapidly.

In order for students to experience the gradual tensing of the vocal cords, students are advised to adopt a forward slumping of the shoulders or a forward motion of the head initially, using very tense hands with the fingers spread out and the palms facing upwards, then tense up their arms and the whole of the upper body with the elbows held close to their body, then gradually push their hands up directly over their heads while pronouncing the Tone 2. This movement was used a great deal when practising counting from one to ten in Mandarin because ten is pronounced in Tone 2 as “shí”. The movement thus reproduces the tonal contour of this tone.
Tone 3: the vocal cords become lax immediately after tense, and then tense up again. However, this description is only accurate on a lexical level. In running speech, Tone 3 is either realised as a lower level tone before a Tone 1, Tone 2, Tone 4 or a neutral tone or a Tone 2 before another Tone 3 according to the tone sandhi rules. Therefore, it is more accurate to describe Tone 3 before Tone 1, Tone 2, Tone 4 or a neutral tone as a lower level tone which requires the body to relax; and as a Tone 2 before another Tone 3 thus requiring the same movement as those described for Tone 2.

In instructing students to produce the low level Tone 3, students were advised to adopt a relaxed, forward slumping of the shoulders accompanied by a forward motion of the head while similar to nodding to produce the sound. In instructing students to produce the Tone 3 before another Tone 3, the movement adopted in producing Tone 2 was recommended with the first Tone 3 syllable.

Tone 4: the vocal cords suddenly tense, and then lax gradually.

When it is necessary to go from tense muscle to lax muscle very quickly such as producing Tone 4, they were instructed to first raise their hands up high like what they were doing in Tone 1, then relax their bodies by bending their heads forward. Furthermore, since in general words with “Tone 4” are louder than the rest, they stamp their foot last.

For instance, in teaching a sentence “wō māma jiào Līlǐ”, students were instructed to push their hands above their head as high as possible with their palms up when pronouncing “māma”. This enabled them to produce the high, tensed sound “mā”. For “jiào”, they were instructed to first raise their hands up high like what they were doing in producing “Māma”, then relax their bodies by bending their heads forward. Furthermore, since the “jiào” and in general “Tone 4”
are louder than the rest, they stamp their foot last. Once these movements were mastered, they were asked to hum the sentences again while walking in a circle. In teaching “wō”, students were instructed to adopt a forward lumping of the shoulders for Tone 3 as the production of this tone requires a relaxed posture. This was again done for 5 times. This constituted step 4.

These gestures here refer to the sets of involuntary movements which vary according to the temperament of the speaker. They stress the message but are not part of an explicit code of communication. They are not the codified gestures which are part of a communication system. Though artificial, they are nevertheless important because they help to set up the overall body tensions for the production of the required speech. Therefore, they seem to be useful in highlighting the importance of tones and other prosodic features in Mandarin.

Corrective gestures such as these set up the necessary body tensions for the production of intonation patterns and for sounds. It is therefore important to also encourage learners to use these gestures in self-assessed study so that they are of assistance in everyday communication.

By the time the sensitisation session had been finished, they had had a considerable amount of intensive practice, in the production of melodic patterns, accompanied by a development of their self-synchrony. In this way, students would have internalised prosodic models not only at the level of the ear but also at the level of the body. When asked to reproduce them, they would be able to “remember” the pattern on the basis of their acoustic impressions and also on the basis of a set of physical tensions which would complement each other very effectively. Together with humming, clapping to the rhythm of the language, movement and gestures, perception was occurring simultaneously through a number of different mechanisms.
The movement and gestures described in this section have much in common with the principles adopted by Brüll (Brüll, 2003).

4.3.1.5 Step 5: Mouthing the words

In this step, the teacher instructs students by saying “Continuing with the movements, now mouth the sentences while I say them out loud.” (Step 5)

For the first time in the learning sequence, so far, students were hearing an intelligible sentence. They were asked not to say anything but merely to mouth the words.

Mouthing the words gives students the opportunity to practise the articulation of the sounds of the words without, in fact, placing them on an intonational background actually produced themselves. This technique should lead to a reduction in the number of articulation errors. However, it is not implied that the prosodic elements of the sentences are no longer being practised. On the contrary, they are very much preserved and actualised, though not vocalised, through gesture.

The approach which isolates each element of articulation e.g. filtered intonation, humming and mouthing before restoring them to a normal context has the further advantage of eliminating as many difficulties as possible in terms of comprehension of the sentence. Consequently, by the time students are actually asked to repeat a full sentence, they will have practised each of its constituent elements many times. They will look forward to achieving success in the next step of the process which will follow naturally and which should present little additional difficulty.
4.3.1.6 Step 6-7: Adding words to the intonation patterns

The teacher then says “Now repeat after me, and then add words to the intonation.” This again is done for five times. (Step 6)

The teacher then instructs each individual to repeat the sentence by themselves; checking that each student is reproducing the sentence correctly (Step 7).

The prosodic patterns are hummed again by the teacher for a further 5 times, and students are asked to say the sentences at the same time as they hear the prosodic patterns. This provides a transition between the kinds of exercises performed so far and the production of normal speech. This step is analogous to singing songs. The humming represents the “tune” to which the students “sing” the words. In this manner, the prosody acts as a very real support to the production of speech making the transition to normal speech a fairly painless process.

4.3.2 Repetition exercises

To further consolidate the gains obtained from the above procedures exercises are repeated with gestures and followed by humming. Repetition first takes the forms of chorus work and then individual repetition. Chorus work provides an environment where anxiety about speaking an L2 could be reduced to a minimum. Although carrying out chorus work in class is “safe”, it is hardly ever likely to be encountered in real life. It is important, therefore, that students also be conditioned to speak with self-assurance in the normal communication situation. To this end, at the appropriate juncture such as at the end of a conversation, a ping pong game could be played with the newly acquired TL. For instance, after practising the following conversation:

A:你好，您贵姓? (Hello, what is your honourable surname?)

B:我姓 Davies, 你呢? (My surname is Davies, and you?)
A: 我姓刘，你叫什么名字？(My surname is Liu. What’s your name?)

B: 我叫玛丽，你呢？(I am called Mary, and you?)

A: 我叫天星. 认识你我很高兴. (I am called Tianxing, it is nice to meet you.

B: 认识你我也很高兴.(It’s nice to meet you too.)

In the ping pong game, a ball made out of a piece of used paper, is thrown around the group while the group is still in a circular formation. The person holding the ping pong ball plays the role of Person A. He/she then throws the ball to another student who plays the role of B after saying, for example, 你好, 您贵姓? (Hello, what’s your honourable surname?). When B catches the ball, he/she needs to answer the question asked with the appropriate answer. In this case, B: 我姓 Davies, 你呢? (My surname is Davies, and you?). After saying “你呢？”, B throws the ball to another student. This game was played after learning every two sentences in the above conversation.

Experience testifies that as the students’ confidence grew, so will their willingness to participate in this activity (Lian, 1980). Though the group situation could not be totally under a student’s control, to some extent, it was possible for students to work at their own pace within the group situation because each time they listen and observe another student’s production of the utterance, their own learning was being reinforced. Through the ping pong game, students get the opportunities to listen to his/her classmates using the learned sentences to communicate thus revising the newly learned language in his/her own head. The randomness of the game also creates a bit of tension which encourages the students to rehearse the language silently while observing other students performance. For the teacher, if a common error occurs, this game gives her/him an opportunity to correct the error using gestures and rhythm.
Repetition exercises such as this obviously provided reinforcement at both the perceptual and articulatory levels. By the time the students completed an “average” sensitisation session, they would have repeated or been exposed to the same pattern or a set of closely related patterns in their situated context about 35-40 times. Such a high number of repetitions were, of course, a great reinforcer of perceptual and articulatory skills.

4.3.2.1. Step 8 Checking for meaning

The teacher now asks “Now what is the meaning of the sentence?” Students volunteer the meaning of the utterances in English and the meaning of the sentence is usually established in seconds. (Step 8).

Notice that throughout the learning sequence, translation and writing down the sentences were not needed. By the time students came to write down the meaning, they had already internalized the tones of the sentences. At the end of the lesson, students were instructed to sit down and write down the meaning or whatever notes they wanted to make for themselves.

In each 2 hour session, every sentence was presented and practised using the above procedure. At the end of each lecture, the whole class engaged in pair or group work in conversation activities using the materials covered in the lesson (Step 9).

4.3.3 Exposure to authentic materials

Many textbooks and traditional teaching methods employ simplified materials. In such a textbook, the writer does not only imagine a sequence by which an imaginary life in the target culture can be conducted, he/she also needs to somehow reduce the language used to conduct this imaginary life based on what is perceived as easy and what is difficult. For instance introducing oneself to complete strangers in a L2 is imagined to be a simple task whereas
asking someone about the location of a hotel is more difficult. The truth is many of the tasks which adult students want to perform within the target culture, at the beginning of their cultural engagement, are much more complicated and fraught with risks than a whole range of tasks they might subsequently wish to perform. In fact, it is not possible in practice to predict the subsequent desires that the student might have with respect to a particular transaction. Thus approaches to the teaching of culture which imagine culture as a discontinuous and heterogeneous assemblage of practices but nevertheless attempt to predict what practices shall be encountered by the students must by and large fail.

Instead it is better to allow students to imagine situations pursuant upon their own perceived goals and desires. However, this is also problematic in that many of the goals and desires, whether perceived or not yet perceived by students, are unpredictable and unknowable. There is, however, a solution to the quandary of selecting what to teach with respect to culture and/or language. The methodology that concerns this thesis responds to this quandary in two ways:

1. Introduce a set of conversations used in authentic relevant contexts and then supplement them with homework tasks involving authentic texts, and other useful web-based materials, students might find that these situations are relevant to their needs.

2. Use an oral assessment procedure in which students were invited to form a pair in constructing and performing an imaginary conversation using the language learned. In inviting students to perform the parts of an imaginary conversation in the target culture, students not only have to produce appropriate language, they are also called upon to respond to any changes in the performance of other participants. In a sense, such situations resemble the kind of authentic situations they are likely to encounter in the target culture.
4.3.4 Choice of teaching materials

One of the difficulties encountered in the teaching of Mandarin tones, in any approach, is the
selection of models to be taught. The problem can be attacked at one of two levels. Ideally, the
student ought to be taught to produce all possible intonations for all possible situations. This
cannot occur for a number of practical reasons not the least of which is that, for the time being
there does not exist an exhaustive description of prosodic patterns in Mandarin. The second is
linked to pedagogic objectives. Is it necessary for students to be able to utter all possible
melodies? In the initial stages of learning to speak a language is it not possible for them to
recognise the function of many patterns while mastering the production of only a few? Studies
in French offer some evidence that this is possible (Lian, 1980). Supporting evidence came
from Brüll’s experiment of using speech-movement as therapy for two boys who were suffering
from apraxia. After a considerable period of training with a limited set of sentences, one of the
boys was able to extend the learned patterns to new words and phrases of his own choosing
(Brüll, 2003).

The teaching materials used in this course for the first year were based on a set of commonly
used situations for travellers. 90% of the vocabulary covered in the course belonged to the
character frequency list provided by the National Standard Chinese Committee (国家汉语水平
考试委员会). While they did not represent “emotional” speech (which has been recommended
as most useful by Verbotonal System of Correct Phonetics) (Renard, 1975), the utility of the
situations and the clarity of the sentences were likely to be more useful and motivating for
students’ internalisation. With time and continued exposure to authentic Mandarin speech
patterns, acting as a kind of foundation, these patterns could then be modified to deal with more
“affective” or expressive situations. Consequently, this course ought to be seen as no more than
a beginning in the study of Mandarin tones, intonation and rhythm. It should be pointed out that even within any one of the selected models a measure of variation is possible. Indeed, for any one sequence of words many prosodic patterns are possible in different emotional contexts due to the effects of stress, intensity and duration of the patterns. For instance, in the sentence:

你喜欢不喜欢吃饺子?

Nǐ xǐhuan5 bù xǐhuan5 chī jiāozi?

Do you like eating dumplings?

Figure 4.1(a) The pitch curve of the sentence “Nǐ xǐhuān bù xǐhuān chī jiāozi?” in which “nǐ” is not stressed.

Figure 4.1(b) The pitch curve of the sentence “Nǐ xǐhuān bù xǐhuān chī jiāozi?” in which “nǐ” is stressed.
169

The stress is on xǐhuān bù xǐhuān not on Nǐ, so the contour for “Nǐ” is shorter. The length of “Nǐ” in this case is 156.5 milliseconds (ms). The stress is on “nǐ”, consequently, “nǐ” is being stressed, resulting in it being longer in duration (531ms) with a very different contour.

This is true for Mandarin as well as other languages. Consequently the model chosen is only one amongst many. This should be made clear to students from the very beginning, as should the fact that sticking to the models suggested would see them through many, if not most, common situations in everyday life. The choice is made for those patterns which represent the essential tensions of Mandarin.

In this course, students were provided with a course data CD-ROM, audio CD and a speech processing tool (Sptool) to facilitate learning. The text based teaching materials consisted of course notes designed according to a number of frequently occurring situations such as “Introduction”, “Talking about one’s family” and so on (see Appendix 2 for the Content page of the course materials). In the course notes, each sentence or vocabulary was recorded separately and linked to a normal version of the sound file and a filtered version of the sound file which is indicated by “F” beside each sentence (see Figure 4.2). Upon clicking the normal version of the sentence, a comparison of the student’s input with the model was immediately available through Sptool (Zhang & Newman, 2003). If the filtered version of the sentence was clicked, students could see the pitch curve but hear the humming rhythm of the sentence as the consonants and vowels had been filtered out. In each chapter of the course notes, conversations utilizing the sentences, relevant vocabulary, relevant grammar points and characters as well as homework tasks were also provided. The homework tasks consisted of conversations or video skits covering similar situations. Students were required to transcribe these video skits or conversations so that they were exposed to a range of conversations spoken by different
speakers. Through these materials, students were trained to recognise some of the expressive patterns in naturally occurring situations as well as get use to the voices of different NSs. The result of such training would be that students would gain mastery of a small but significant number of patterns for real communication while learning to interpret a much broader range of patterns as well as internalizing the essential tensions of Mandarin.

Figure 4.2 Sample page of the teaching material

4.3.5. Selection of Vocabulary

The vocabulary used in the teaching materials was based on the most commonly used Mandarin words according to the frequency list published by the National Standard Chinese committee. They had been regrouped according to their most common associations with each other within
any particular theme. A selection on such a basis had two advantages: (i) if students already knew the words, then the sentences would have a reinforcing effect; (ii) if this was not the case, then it would be one of the ways in which students were introduced to essential vocabulary which was already partially contextualised by virtue of being placed in its most common linguistic environment.

In the classroom procedure described above, the sentences consisted essentially of repetition of sentences taken from their lecture notes. These sentences were usually presented in a context that is appropriate for the language. Students were also required to do extra tasks involving similar contexts such as homework. These extra tasks were provided on the subject data CD-ROM that accompanied the course materials (see Appendix 1 for homework examples at the end of Week 1’s work).

The smallest unit of the language being presented was a sentence rather than individual words or compounds. This was because in Mandarin, the acoustic characteristics of the words change when they are in a sentential environment. For instance, when a word is read in isolation, the tone of the word is different from when the word is part of a sentence (Guo 1991). So concentrating one’s effort in mastering the tones of individual words or compounds does not guarantee success in producing the sentences containing those words.

4.3.6 Use the text only as a last resort

The sentences taught in the teaching materials were presented aurally first during the sensitization session and not in written form. The elimination of the Pinyin romanization from the teaching material meant that the written character texts could not be read (unless the students looked up each character in a dictionary) and had to be heard first. However, students’
urge to write down the pronunciation using their own spelling was too great. To delay this urge to romanise, steps 1 to 7 were carried out as described first. By following the first seven steps, students’ urge to write things down was realised as a reinforcement of the TL based on the correct perception of the TL rather than as a barrier to perception. If students were allowed to write down their own romanised spelling through listening to the sound files first, students would inevitably use the written spelling for recalling the language. By introducing the language through movement, gesture and so on, students were forced to rely on their memory and senses to recall the TL learned at least from step 2 to step 7. In truth, if they were allowed to write the spelling down first, the energy spent in steps 2 to 7, no matter how well meant, would probably have been wasted.

4.3.7 Role of the teacher

The teacher’s role was absolutely central. Although, to some extent, it may be possible to use this methodology for self-teaching purposes, the students’ “deafness” may well prove to be a very difficult barrier to break down unaided.

4.3.7.1 Supervision of the sensitization session

For the best possible results, therefore, a teacher must supervise at least the sensitisation session. It is in this session that he/she would ensure that the prosodic structures have been perceived correctly. If incorrect perception of the prosodic structures has been detected, he/she must take immediate corrective action. Moreover, he/she must also ensure that students were developing the necessary “feel” for the language and self-synchrony (refer to 4.3.1.4. Step 4: Incorporation of movement and gesture).
4.3.7.2 Monitoring of learner performances

It was essential that the teacher should constantly monitor all the students’ performances. It has been argued that a teacher working under such conditions would be incapable of picking out the errors committed by students, because he/she cannot listen to all of them individually. Experience in this course indicated that in the majority of cases only one or two persons in a group would be significantly “out of tune” with the rest of the group. Such persons were easily identified and corrected immediately provided that the size of the group was no more than 30 students per group. In this way, students did not spend a considerable time reinforcing their errors during “private” repetitions as they were likely to do in a language laboratory.

The teacher was also in a position to determine whether generalized errors are occurring and to correct them in a variety of ways e.g. by exaggerating the model sentences at the point of difficulty or by using corrective gestures. In addition, when monitoring chorus work, students need not rely wholly upon their auditory perception in order to locate and analyse errors. For example, students would be able to see if body tensing movements occurred when they should not or whether self-synchrony was lacking.

4.3.7.3 Establishing links to real face-to-face communication

The teacher should act as a kind of buffer to the artificiality of the learning situation. Students needed to be told that whatever learning situations they were in, such situations would necessarily be artificial to some extent and different from those to be encountered in real face to face communication. In order to reduce the gap between real communication and exercises of this nature, the teacher should attempt to establish the kind of situations in which the sentences might be uttered e.g. asking students to maintain eye contact (rather than closing one’s eyes in an attempt to recall words) when speaking to each other. He/she should also remind students of
the importance of gesture in the communicative act and should demonstrate this by taking part in the activities of the group.

VTM in and of itself is sensitive to the variations required by the diversity of real social situations. In fact, the teaching techniques described in Steps 1-8 teach more than the few phrases chosen each week. In teaching the rhythm, and the eurhythmic gestures, we are also imparting onto the students how Mandarin speakers synchronise with each other in speech, in proximity and in body language.

In the course under study, students were strongly encouraged to form conversation exchange groups with NSs in the community so that they could put into practice the prosodic patterns they had been learning. This was the privileged place for the actualisation of prosodic patterns and for hypothesis testing with real NSs of the TL accompanied with the realities of communication.

4.3.7.4 Teaching learners how to learn

Apart from employing a pedagogical sound approach, another important factor in this methodology is the teacher’s rapport with the students. He/she must be able to establish a good rapport with the class as a whole and with the students as individuals. Throughout the teaching/learning process, the teacher should also inform students that though the teaching method is fun, they would still be required to do a significant amount of work. This implies that he/she must possess some knowledge of the elements of VTM, articulatory phonetics and kinesics. He/she must also be able to develop the ability to diagnose errors in tones, intonation and in synchrony. In other words, they must also be able to pronounce Mandarin sounds properly. Last but not least, the teacher needs to be able to sit back and encourage students to
produce the language. In the case of Mandarin, language correction should move away from the traditional verbal correction such as “no, it’s a third tone, not a first tone” to include provision of feedback through movement and gesture in the appropriate prosodic contexts.

In summary, in the learning sequence described, several rather “unconventional” elements make an appearance. For instance, relaxation exercise, humming, body movement and gestures, mouthing the words and then repetition are all present. The nine steps of the learning sequence offer students a range of physical ways for recalling tones. These measures set up a series of learning steps that can be used for self-access learning at home. However, what other sort of feedback can we give students in their self-access private study?

4.4 Computer enhanced learning support

4.4.1 Teaching materials

Having been sensitised in both perception and articulation, students are now in a position to make efficient use of the computer materials provided for the course. They were provided with a course data CD, audioCD and Sptool to facilitate learning.

The text based teaching materials consisted of course notes designed according to a number of frequently occurring situations such as “Introduction”, “Talking about one’s family” and so on (see Appendix 2 for the Content page of the course materials). In the course notes, each sentence or vocabulary was recorded separately and linked to a normal version of the sound file and a filtered version of the sound file which is indicated by “F” beside each sentence (see Figure 4.2). Upon clicking the normal version of the sentence, a comparison of the student’s input with the model was immediately available through Sptool. If the filtered version of the sentence is clicked, students can see the pitch curve and hear the humming rhythm of the sentence as the consonants and vowels have been filtered out. In each chapter of the course
notes, conversations utilizing the sentences, relevant vocabulary, relevant grammar points and characters as well as homework tasks are also provided (see Appendix 1 for a sample unit). The homework tasks consist of conversations or video skits covering similar situations. Students were required to complete the homework tasks so that they were exposed to a range of conversations spoken by different speakers. Through these materials, students were trained to recognise some of the expressive patterns in naturally occurring situations as well as to get use to the voices of different NSs. The result of such training would be that students would gain mastery of a small but significant number of patterns for real communication while learning to interpret a much broader range of patterns as well as internalizing the essential tensions of Mandarin.

In the course notes, each sentence or vocabulary was recorded separately and a comparison of the student’s input with the model was immediately available through Sptool. This provision of immediate access to feedback ensured (i) that both the perceptual and articulatory patterns were preserved as much as possible in the students’ memories at the time of comparison and (ii) that their attention was still centred specifically upon the utterance which they had just produced. A filtered version of the same sentence or vocabulary was also provided. This version is indicated by “F” in the material.

4.4.2 Function of Sptool

In the present study instead of adopting an ASR-based CAPT system to enhance learning and expect such a system to provide accurate and informative feedback to beginning learners of Mandarin, Sptool (Zhang & Newman, 2003) for tone languages was developed and tested with the experimental group of students. It was designed to give learners the opportunity to listen and perceive differently, thus increasing the probability of changing the ways they both hear and
produce. The decision to create a speech tool specifically for tonal languages was made because the existing speech tools such as Winpitch (Martin, 2003) and Praat (Boersma & Weenink, 2003) are difficult for learners to use as they were initially created by phoneticians for phonetic analysis. By contrast to other ASR-based CAPT system, Sptool (Zhang & Newman, 2003) is designed to offer feedback that is non-judgmental, i.e. without a scoring system and allow students to explore and reflect during the process of learning, and not just at the end of the learning process. When using the Sptool, students can listen to the language and then reflect on it by observing visually the differences between their production and the NS model.

All the sentences in the course were linked to their corresponding sound files (links are in blue on the dataCD but are underlined in the sample learning materials in Appendix 1) and were linked to the Sptool (Figure 4.3). Once clicked, the Sptool would open and the learner could listen to the teacher’s model pronunciation by clicking on the “teacher” icon. With one click, he/she could hear the model sentence and see the pitch curve of the model sentence on the screen. The vertical axis (Y-axis) represents frequency and the horizontal axis (X-axis) represents duration (measured in milliseconds). If the learner wanted to hear a smaller chunk of the sentence, then he/she could select a portion of the curve by dragging the cursor over the part they wanted to hear. After listening to the sentence numerous times, the learner could decide whether he/she wanted to record their own production.
In Figure 4.3, the blue line represents the NS model and the green line represents the student’s rendition.

Figure 4.3 A screenshot of the speech analysis tool

Before clicking on the record button, however, it was necessary for the students to tell the program whether he/she was a female or male student. This was necessary because due to anatomical differences, females generally have higher fundamental frequencies than males. The difference is sometimes as much as over 100 Hz. For instance, if the program was configured to measure a female voice, the pitch curve of a male learner would not be displayed correctly. However, once configured correctly, the pitch curve of the learner’s recording would then be displayed properly. The needs of high pitched males and females were also looked after by setting the male voice range between 60-250 Hz and female voice range between 100-390 Hz. These voice ranges could be changed by clicking on the Edit button and then click on “Option”. Once “Option” page (Figure 4.4) had been clicked, the following screen would be displayed:
On the “Option” page, the male and female voice ranges could be changed to suit a particular student. Furthermore, the voicing threshold could be reset to measure more of the signal in the audio file. For instance, if the voicing threshold is set to 0.04, then only the voiced component of the sound signals (words) is measured and displayed. However, if one desires to measure more of the signal, then the threshold need to be set at 0.02 or lower. At this level, both voiced and weaker signals can be measured. The speech threshold indicates the point where sounds are actually picked up. The framestep measured how many samples the program is going to use to measure the sound signals. The higher the framestep value the more detail the sample can be measured.

The course dataCD contained teaching materials in html format; all the associated sound files, Sptool and short video skits which students could test out their comprehension of the new
language learned through watching them. An audioCD was also provided with the course materials. Students taught with SEA in 2003-4 also undertook a weekly compulsory class using computer-enhanced teaching materials. Students could also carry out such tasks at home.

4.4.3 Role of Sptool

Steps 2 to 7 in the learning sequence were duplicated in different forms through the use of the Sptool. While the classroom sequence was more or less teacher-driven and active, the Sptool allowed the gesture and movement procedure to be experienced in a different way.

Figure 4.5 Picture of the Sptool showing the sample sentence: “Hello, what’s your honourable surname?” (highlighted portion in white)

The beat, stress, word groupings, key words, and sentential intonation are all indicated in the pitch curve and the sound file. In the above sample sentence in Figure 4.5, nín guìxìng (您贵姓, what’s your honourable surname?) is a string of key words (highlighted in the wave form) being displayed and the curve clearly showed that while both “guì:honourable” and “xìng: to be surnamed (verb)” are both in Tone 4, different emphases had been placed on them resulting in “guì” being longer in duration thus more emphasized. This illustrates the fact in running speech
of Mandarin, that is, not every syllable or tone is spoken in the same manner. Students also need to pick up the stress cue which is indicated by the loudness of “gui” in this instance. In other words, learning Mandarin using Sptool, students gain access to information which cannot be adequately described in words.

In general, the height (related to the muscular tenseness of the body) is indicated clearly with respect to other tones. The height of Tone 1 and Tone 4 remind the students of the need to stretch their voice range beyond their normal voice range. This information was very useful in enabling students to physically change their way of producing the target sentences after observing the differences between the NS’s production and theirs. The changes could be assisted by the movement and gestures taught in the sensitization sessions.

Students could also select any portion of the sentence for listening practise and repetition. For instance, in the sentence as shown in Figure 4.6,

Wǒ shēngyú yī jiǔ wǔ líng nián bā yuè èr shí sān hào.
我 生 于 一 九 五 零 年 八 月 二 十 三 号。
I was born on the 23rd of August, 1950.
Figure 4.6 Pitch curve of “wŏ shēngyú yījiū wǔ lǐng nián bā yuè èr shí sān hào”
( 我 生 于 一 九 五 零 年 八 月 二 十 三 号。)

The production of “shēngyú” requires the body to be tensed and kept tensed in order to produce the next “yī: one”. Students could select the three syllables “shēngyúyī” (the highlighted portion of the wave form) in order to explore how physically one had to keep one’s body tensed to produce this group of words. In authentic Mandarin speech, these three syllables do not form a syntactic group. So logically, this sequence of words would most probably not be chosen by a teacher as a practice item for the students. However, students would have or did have difficulty in pronouncing these words together in reality. Therefore by allowing students to practise at their own pace and make their own choices, they are able to work on what they deem to be important for themselves through the flexibility provided by Sptool.

Many of the things that could be done using the program may not have been initiated or thought of by teachers but were being spontaneously explored by the students through use. Furthermore, being able to experience each sentence repeatedly through the Sptool created an environment in which students could spend more time on the language through listening to them repeatedly.
Another important contribution of CAPT software similar to Sptool is that in using the tool, students inevitably go through the sequence of intense listening to the model, imitating and then finally producing the language. This way, students become increasingly familiar with range of language patterns in a L2. In this study, learners did not go through a period of training perception as advocated by so many researchers (Gary, 1975; Postovsky, 1974). Although it is likely that student learning would be further improved if discrimination exercises did accompany the sentences.

Together with the provision of the teaching materials on CDs, Sptool improved students’ access to the TL on a massive scale as students had easy access to the TL and feedback support at the click of a mouse.

4.5 Feedback provision

It may have been possible to support learners in the program by providing them with individualised corrective program. Such a program could have been determined by looking at the results of students’ mid-semester oral examination and other tests. Such diagnostic program for each individual student would be extremely time-consuming and thus was not attempted in the present study.

4.6. Individualization of the curriculum

In this thesis, the act of language learning is treated as no different from the ways human beings make sense of the world through engaging with complex external multi-sensorial information. As such, language learning involves learning to cope with needs that are not always categorizable or understood. Learners also often have to cope and work with the unknown, the invisible, the unpredicted and the unpredictable.
Creating a learning system that is geared towards catering for the individualistic and unknown needs of a large number of students and the other unknowns they might need to deal with albeit they remain unknown or invisible means that individualization of the curriculum is inevitable and desirable. Such individualised curriculum needs to activate as much in the learning subject as possible.

4.6.1 Connection and Flows

Two conceptual tools from the Deleuzian toolbox are used in the development. The first one is “connections and flows” (C&F) (Deleuze & Guattari, 1987) and the second is the concept of “rhizome”. Both concepts re-examine what constitute good practices in learning. For instance, in the field of language learning, good connections which enable a flow of information resulting in learning are usually defined categorically by linguistics or applied linguistics. In other words, if we want to enhance language learning, we must choose connection points which are necessarily linguistic. For instance, contrastive analysis of L1 and L2 might enable teachers to impart a list of errors that L2 students need to avoid. Good practices in language teaching are those based on research findings from contrastive linguistics.

The Deleuzian conceptual tools of “connections and flows” and “rhizome” (Deleuze & Guattari, 1987) question such categorization of good or bad practices or behaviour according to categorical means. They argue that if good C&Fs are opened up to things outside those legitimised by linguists, the process of learning could be vastly enhanced from the point of view of the learners. For instance, in the SEA learning environment, the adoption of the practices from VTM allows connections and flows to be formed between learners’ physical bodies and the L2 they are learning. This is a far cry from the practices which deem the mastery of grammar as the only way to learn an L2. In VTM, authentic materials were played through very
loud public address systems so that the inflections and modulations of native tones were experienced in the tissues, of the stomach, chest, arm, and legs. In the language learning environment that is the subject of this thesis, practices derived from VTM have been used to create redundancies and a layering of memory traces which cognitive psychologists and applied linguists agree to be amenable to the acquisition of L2.

4.6.2 The Rhizome

The second conceptual tool from Deleuze and Guattari is the rhizome structure. The rhizome structure has been adopted in this thesis because it allows indifferent categories to connect and achieve learning. The rhizome structure (Deleuze & Guattari, 1987) has certain traits and some of the traits are:

1. Total connectivity: each element in the rhizome connects up to every other element.
2. Heterogeneity: the elements that make up the rhizome are of different sorts (in fact all possible sorts. These include linguistic things such as grammar, phonemes, sentences but also non-linguistic things such as a tune from one’s childhood, a cartoon character, a smell, a touch, a sensory experience from one’s past).
3. Multiplicity: the rhizome cannot have a centre or core and cannot be thought of in terms of categories without losing its rhizome-ness. In other words, no one meaning is privileged over another. Words can have a multiplicity of meanings.
4. Asignifying “ruptures”: The rhizome is discontinuous and these discontinuities do not necessarily “mean” anything. In fact the rhizome means nothing in itself. There is no truth that transcends or traverses the rhizome. All truths are merely elements in the rhizome.

In other words, the rhizome can be imagined as a thing, made up of all different sorts of things, in a constant state of change and flux.
In SEA, students can bring any materials they like to the learning environment and arrange them any way they wish, according to whatever self-determined strategies and projects they currently have under development. Students' desire is not "repressed" but encouraged and stimulated via the provision, in terms of environmental design, of a wide range of different sorts of practices, tasks, apparatus and materials. Thus, many practices of SEA harness, and place in consort, a range of phenomena and behaviours that are generally seen as disparate and "not belonging together". This is not a theoretical weakness of SEA but a theoretically informed and considered approach.

4.7 Self-access learning

One of the ways in which curriculum design can be approach in a more rhizomatic manner is through the incorporation of the so-called self-access learning. Self-access learning here refers to any sort of self-motivated learning which do not usually happen in a language or computer laboratory. In this study as many students were mature-age students, they only had time to study at home. In the design of the learning environment, self-access learning was encouraged and catered for through the provision of course materials on data and audioCDs. The dataCD also contained homework tasks that were designed for L2 learners to further explore Mandarin. The self-access learning tasks were designed to be exploratory rather than directed. Tasks typically included transcription of short Mandarin sound files or video skits, or essays on the materials covered. Each task had feedback embedded in it so that each step of the discovery process would give learners a sense of achievement. Even in the assessment phase of the learning, students negotiated tasks in self-selected groups for assessment (see section 5.6, Chapter 5). This was preferred precisely because student -created conversations tended to open up questions
where the answer was not already pre-ordained even when a limited range of language was covered.

4.8 Minimum use of translation exercises and discussion of grammar

The study under investigation in this thesis suggests that conscious attention to tones (e.g. via diacritics) is as likely to be a major source of error as it is conducive to accuracy. Similarly translation exercises also call for the conscious recall of the TL by comparing the two different grammatical systems. This kind of activity is likely to cause more errors to be made in spontaneous speech making due to the differences in the grammatical structures of the students’ mother tongue and TL. In teaching students to speak a L2, students have always been advised to “think” in the TL. Students trained with SEA will be much more capable of “thinking” in Mandarin while speaking than students trained without the context of translation exercises. This is the reason why in SEA, translation is seldom used as a learning activity.

4.9 Summary

This chapter is concerned with the description of the learning environment that enacts the pedagogical model of this thesis. The justification for the various methodological principles adopted in the learning environment under study has also been outlined. Evidence from language acquisition research arguing against the validity of the critical age hypothesis offers researchers hope that what is being investigated in this thesis is not a waste of time.

Discussions on the use of Pinyin and tone diacritics clarified the position that is taken in this thesis on the necessity of reducing the processing load on our L2 adult students during the initial phase of Mandarin acquisition. As both Pinyin and tone diacritics are artificial representational systems of features of Mandarin defined according to some insufficient categorical systems, the
inclusion of both would only discourage learners from the *direct perceptual experience* of the TL (Steiner, 1996). In removing Pinyin and tone diacritics, we are actually removing two sources of artificial representation. In our model, applying Newport’s “less is more” principle (Newport, 1990), we have effectively taken steps to enhance students’ aural experience of the TL speech stream and allowed them to process fewer sources of information than usual.

Due to the complex changes that occur when lexical tones are in the environment of other tones in sentences, the decision to teach in sentential contexts is aimed at sensitizing students to these changes and their lexical counterparts. The emphasis on the use of authentic materials with practices derived from VTM further draw students’ attention to the suprasegmental elements of Mandarin without impoverishing the actual experience. Similarly, the decision not to discuss grammar nor use translation as a means of learning is also designed to reduce the processing load on students especially when they are called upon to produce the TL covered in speaking. It is also designed to reduce the risk of confusion by not subjecting exclusively learners to pre-established explanations.

However, it would be erroneous to think that the methodological principles adopted in this thesis are designed to close down learners’ options. The highlighted principles have proved to benefit the initial stages of audition in learning a FL and therefore they deserve to be introduced and practiced as possible tools for future learning. Thus, the methodological steps described in this thesis should not be seen as prescriptive but thought of as possible paths a L2 learner could take in the process of “becoming” a proficient Mandarin speaker.

In the next chapter, the design of two experiments involving the experimental and control groups of students will be discussed.
Chapter 5 The study

5.1 Subjects in the study

The subject entitled Chinese Language and Culture Part 1 was read by beginning students of Chinese at the University of Canberra in 2001, 2002, 2003 and 2004. Students from 2001 and 2002 constitute the control group. Students from 2003 and 2004 constitute the experimental group. Students from both groups were evaluated with respect to their performance in producing intelligible spoken Mandarin after the first six weeks of the course. All of the students speak English as their L1 and were total beginners of Mandarin. Students in the control group (CG) were not taught with SEA and did not have the benefit of the interactive computer support. Students in the experimental group (EG), on the other hand, was taught by SEA and had available to them a range of interactive computer programs including a speech processing tool which they could use at their own leisure to listen to course materials spoken by NSs of Chinese. This speech processing tool (Sptool) (Zhang & Newman, 2003) displays waveforms and $F_0$ contours and allows students to record their own rendition of the materials and compare the corresponding $F_0$ contours.

Two experiments were conducted in order to analyse the differences between the spoken Mandarin conversations from the experimental and control groups. The first experiment determined the differences between the male and female speakers of the two groups in the range of their fundamental frequency. Fundamental frequency ($F_0$) is a term derived from the study of the physics of sound, and used in acoustic phonetics, referring to the lowest frequency component in a complex sound wave (other components being known as the “harmonics”). Frequency refers to the number of complete cycles (opening and closing movements) of vocal cord vibration in the unit of time (per second). The “fundamental”, or $F_0$ (“f nought”), is of
particular importance in studies of intonation, where it displays a reasonably close correspondence with the pitch movements involved (Crystal, 1997).

The second experiment involved a perceptual experiment in which the oral performances of both groups were judged randomly by nine female NSs of Mandarin. The researcher was not one of the NS markers. This was used to obtain from the subjective ratings of the NSs’ on the overall quality of Chinese-L2 production of the two groups. In addition, differences between the groups on the prevalent tonal errors were analysed through an auditory analysis with three experienced teachers of Mandarin as judges.

Both quantitative and qualitative data were collected. The quantitative data collected consisted of oral performances in the form of spontaneous speech. Qualitative data collected consisted of an end of semester questionnaire on the use of the computer enhanced learning resources and general work patterns, and interview data collected from students through face-to-face interviews with the researcher.

5.2 Methods of data collection

5.2.1 Choosing the appropriate testing materials

Choosing the appropriate testing materials for a study such as this one had not been easy in the field of teaching Mandarin as a second or FL. Many researchers have commented on the effect of different testing materials on the outcome of research. For instance, in Miracle’s study (1989), students were presented with sentences in Pinyin romanization with tonal diacritics and English translation. Miracle (1989) argued that tonal diacritics were used so that the students would not have to resort to guessing the proper tone of words which they may have forgotten. In Xiaonan Susan Shen’s study (1989), the subjects were asked to read a familiar lesson from a textbook written in Pinyin rather than characters as a pre-test experiment showed that learners’
attention were diverted to the recognition of the characters instead of their pronunciation. Guo (1991) also provided some evidence of tonal errors based on materials with diacritics, and Pinyin and materials without Pinyin and diacritics. Predictably, Guo’s results also showed that students performed worst using materials without Pinyin and diacritics. The task used in collecting data in all the above studies (Miracle, 1989; Shen, 1989; and Guo, 1991) was reading.

Only one study used spontaneous speech as the data collection mechanism (Chen, 1997). However, the subjects in Chen’s study were not total beginners of Mandarin. Almost all of them had various informal exposure to spoken Mandarin. Though his study yielded a number of interesting findings, the conclusions on L2 learner errors in spontaneous spoken Mandarin is far from conclusive.

The data being analysed in the present study consisted of spoken materials produced by students in both the control (2001 and 2002) and experimental groups (2003 and 2004) using language covered in the first 6 weeks of their Mandarin study (see Appendix 3). Data collected were mainly in spontaneous dialogue form. As the whole approach of this teaching methodology was based on learning and teaching Mandarin in situated contexts without cutting up sentences into isolated words or smaller chunks, it seemed appropriate that the testing format also followed a similar format. Wang and Spence (1999) found that L2 learners of Mandarin can produce better tones with isolated words than with sentences. Hsieh’s (1996) study also found a similar trend with the acquisition of tones in Mandarin (Hsieh, 1996). However, the dialogue was chosen as the primary vehicle for the oral tests in this study because dialogues is the format through which L2 learners in Mandarin are likely to interact with NS in reality. Therefore, insights gained from this study with regards to the L2 learners’ performance in dialogic situations would most likely be more useful in real-life situation.
Spoken conversations from an oral test were collected after 30 hours of face-to-face class contact. This data is used for analysis. For the oral test, students were instructed to form pairs and carry out a list of designated tasks. They were instructed to write the conversations, get them checked by NSs but not the principal examiner (i.e. the researcher) and perform them on the day of the examination. The students were informed that one of the primary aims of the test was to see whether they could be understood by the researcher who was also the examiner for the oral examination without the examiner having prior knowledge of the content of their conversations. This meant that if the students’ dialogues were comprehensible, then the students had achieved the first most important aim of using the language that is to communicate successfully in Mandarin.

Within their oral performances, students were encouraged to include the questions and conversation materials taught during the first six weeks of studying Mandarin. In the first six weeks, a total of 104 sentences were taught (see Appendix 3). In the first six weeks, the materials covered basic language from greetings and introduction, to questions about one’s family and oneself and language needed for a visit to a person’s home. Students were encouraged to be as creative as possible. In other words, they were encouraged to include language not covered in the teaching materials. Students were not allowed to read a script either in English translation or in characters. Therefore, the oral performances were performed totally from memory so that students’ pronunciation would not be affected by the need to recognise characters or influenced by Pinyin.
5.2.2 Instrumentation and research procedures

5.2.2.1 Experiment one: The acoustic experiment

Two experiments were carried out. In Experiment one, 5 male and 5 female students from the control group were involved in the experiment. Similarly, in the experimental group, 8 male and 4 female students were involved in Experiment one. These students all speak English as their L1 and were total beginners of Mandarin. Both Mandarin and English conversational data were collected.

Unfortunately, the lack of Mandarin NSs in Canberra Australia made the collecting of equivalent data (equivalent to the type of data produced by students) impossible. Thus Mandarin speech data from 13 male and 7 female NSs of Mandarin were taken from news broadcasts and other conversation materials. In selecting the data, the influence of discourse on the language produced was not controlled.

From the control group, 10 students’ oral performance data collected at the end of six weeks (30 hours) of face-to-face contact was analysed through both auditory and acoustic analyses with specific regards to tonal errors. Similarly, in the experimental group of students, the data from 8 male and 4 female students was analysed through both auditory and acoustic analyses. The auditory analysis involved three NS listeners, who were also experienced instructors of Mandarin as L2, to judge the tonal production. They separately marked, according to their judgment, the deviant tonal production of each syllable with Chao’s tone letter system (Chao, 1930). They were allowed to listen to each utterance as often as they wanted. A tone was considered finally mispronounced only when judged as inaccurate by at least 2 listeners. Based on the auditory analysis, an acoustic study was carried out to verify the mispronounced tones. The tonal productions of 4 subjects from the control and 5 subjects from the experimental group
were analysed by Praat (Boersma & Weenink, 2003) to extract the starting-F$_0$ and the ending-F$_0$ of mispronounced tones according to their tonal category (except for the neutral tone 5). The mean pitch values of Tones 1, 2 and 4 were computed respectively. Due to Tone 3’s complicated tone sandhi rules, error rate with Tone 3 is expected to be high and the underlying causes difficult to trace. For instance, errors in Tone 3 might be produced because of the complicated tone sandhi rules but the position of Tone 3 in the sentence might also be responsible for the errors. Without a more thoroughly thought out test, possibly using different and more stringent elicitation method, it will be difficult to ascertain the causes of errors for Tone 3. A more targeted experiment can be carried out in which Tone 3 is given a special treatment and then compare the production of Tone 3 of the treatment group with that of the non-treatment group.

The F$_0$ of the voices of the subjects which is measured in terms of the overall mean F$_0$ as well as the mean pitch and pitch range of the speech samples of the 22 subjects had also been extracted using Praat. Limitation of space in this chapter does not allow me to outline in detail the oral performance profile of each student. Nevertheless, an attempt has been made to identify patterns of errors from these students’ profiles. The quantitative results obtained from both the experimental and control groups of students are dealt with in this chapter.

5.2.2.2 Experiment two: The perceptual experiment

In Experimental Two, a listening test involving native Mandarin markers was carried out. The markers consisted of 9 female NSs of Mandarin from the University of Canberra. The researcher was not one of the native speaking markers. All markers participated in the current listening test on a voluntary basis. Research materials were the spoken conversations from both the control and experimental groups of English speaking students. The markers did not know
which group each student belonged to at the time of marking. The number of samples marked was 22.

Markers were given the following instructions:

1. Please listen to the recordings and then assess each of the speaker’s performance on a scale from
1 to 9.

1: being totally non-understandable; 9 being completely understandable and resembling a near NS level.

If a speaker sounds somewhere between highly natural and highly unnatural, circle the appropriate number on the scale. Do not hesitate to use the ends of the scale (1 or 9) when appropriate.

Make sure your rating is based on how close to NS level and how natural the language is to you.

2. Each utterance or conversation can be listened to as many times as required to reach an accurate assessment.

5.2.3 Recording

A set up using Sony microphone plugged into a Dell laptop computer using Cooledit 2000 was used to make the recording. The sampling rate was chosen at 44100Hz with mono channel at 16
bit. The recorded material was later analysed using a speech analysis tool called Praat (Boersma & Weenink, 2003). Using Praat, it is possible to obtain a graphic representation of the waveform, the amplitude and the fundamental frequency of the speech segment. The F0 could then be sampled at any point along the resultant graph. In order to minimize the effect of different length speech sample on the F0 obtained, a 60 second length sample has been chosen from every subject’s sample for analysis. The results discussed in the next chapter are outcome of a closer acoustic analysis and interpretation of the F0 graphs produced in the above manner.

5.3 Typology of errors

One of the first researchers to suggest that learners’ errors were invaluable tools in the study of the language learning process was Pit Corder (1967). L1 interference has long been considered a major cause of pronunciation errors in the learning of a L2. Nunan (1991) indicates that the influence of the L1 is much greater in acquisition of phonology than in acquisition of morphology or syntax. Therefore, it is important to gain an understanding of the source of these errors that are under study. The following is a summary of sources of errors that are thought to be the main causes of pronunciation errors especially in the area of tonal production in Mandarin.

These errors are classified according to various categories established by different theories about sources of error in SLA. These include:

1. Interference errors (Corder 1967);

2. Natural phonology (Stenson, 1975);

3. Induced errors (Hatch, 1983).
These categories are reviewed briefly below. Although they were not developed specifically from the study of tone acquisition of a tone language, they have been associated with, or developed for the study of different aspects of the phonology of SLA. In the present work, these categories have been expanded to cover the study of Mandarin tones. They will be used to evaluate the data collected.

5.3.1 Interference errors

It is generally agreed among researchers that, of all aspects of the language, pronunciation is the one most strongly influenced by the learner’s L1. According to Ellis (1999), the influence of the L1 will occur in all aspects of the L2 but is likely to be most evident in L2 phonology. Ellis further argues that the influence of the L1 is more active in elementary students than in intermediate students, i.e. more active in the earlier stages of the SLA. He also argues that interference is stronger in formal classroom FL learning than in “naturalistic SLA”. These two points are particularly important for the present study as the subjects are beginning students studying Mandarin in a FL environment. This implies that they are undertaking their language study in a situation where L1 interference is likely to be a major factor.

5.3.2 Errors due to natural phonology

Schane (1972) argues that some sound segments in the L2 are more “natural” than others (either because they are more universal or simpler in articulatory effort) and thus are easier to acquire. As a consequence, learners of the same L2 with different L1 s will exhibit common acquisition orders and developmental sequences. Such similar errors committed by L2 learners, regardless of their LI, are known as developmental errors. Though tentative arguments in L1 tone acquisition had been proposed by various researchers (see Chapter 3), no concrete evidence has been available from the research conducted in the L2 learning especially with respect to the
acquisition of tones in a tonal language. More extensive research has to be carried out with learners of the same tonal L2 with different L1s in order to argue for the existence of a developmental sequence in tone acquisition. The aim of this study is not to search for such a developmental sequence but information on the distribution of tonal errors will be provided.

Data often show that some errors made by L2 learners closely resemble those made by children when they are learning their L1. Some researchers suggest that the order of acquisition of phonological segments for adults in L2 learning is similar to that of children in L1 learning. Hatch (1983:22) terms this phenomenon “reactivated L1 processes”. In terms of Mandarin tone acquisition, the natural phonology theory would argue that certain tones are intrinsically more difficult to produce than others and that this will dictate the order in which the tones are acquired. However, the evidence reviewed in Chapter 3, section 3.1.1 is not conclusive. The design of the learning material used in this experiment was not based on the principle of introducing learners to certain “easier” and therefore more “natural” tone combinations. Therefore, evidence supporting the intrinsic difficulty of producing certain tones is unlikely to be easily analysed. Within the scope of this thesis, a search for such evidence was not attempted.

5.3.3 Induced errors

Induced errors are errors made by learners as a consequence of the type of instruction they receive (Stenson, 1975). An example might be where inadequate explanation of two linguistic items causes them to become confused in the mind of the L2 learner. The literature on “induced errors” usually refers to grammatical errors. However, it could be extended, in the case of the teaching of Mandarin tones to cover errors caused by the manner in which the tones are taught. This is particularly relevant to the treatment of Tone 3 in Mandarin. In this study, the principle
of teaching of citation tones in isolation or extensive discussion of tones and tone diacritics was not used. However, the patterns of errors in students’ output might provide some evidence that the treatment of Tone 3 and the associated tone sandhi systems could benefit from some explicit teaching.

5.4 Intelligibility standard

Intelligibility has been variously defined by many researchers in the field of SLA. Kenworthy (1987) sees intelligibility as being understood by a listener at a given time in a given situation. This means that the process of intelligibility entails that the more words a listener is able to identify accurately when said by a particular speaker, the more intelligible the speaker is. Smith (1992) thinks that in order to make sense of the term “intelligibility” one needs to draw a clear-cut distinction between intelligibility, comprehensibility and interpretability. Smith (1992) defines intelligibility as word/utterance recognition; comprehensibility as word/utterance meaning; and interpretability as meaning behind word/utterance.

He argues that the three categories, intelligibility, comprehensibility and interpretability, constitute three degrees of understanding in a continuum: from intelligibility to interpretability in order of importance. Unfortunately, the distinctions between the three categories seem not to be so clear. In the process of testing intelligibility, it is hard to see where one category ends and another one begins on Smith’s categorization. Clearly it is not possible to be intelligible but not comprehensible. Secondly, with familiar topics, the effect of context and other clues enable a listener to decode the speaker’s message more easily. In the case of the testing process used in this study, since students prepared the conversation themselves, the meaning of the conversation is clearly comprehensible to both students and the researcher because students only had access to the language used covered in class in a FL learning situation. Their conversations might even
be comprehensible to their teacher who is familiar with the language they had covered but whether they are comprehensible to NSs is another matter.

Though Smith (1992) defined intelligibility as ability to decipher word and utterances, it is not clear whether this refers to the segmental phonemes only or whether the suprasegmental features of speech also figure in this equation. This is a relevant concern especially in terms of learning to speak Mandarin. If one wishes to gain the respect and admiration of the TL communities, one aspect of mastery must therefore be the control of the suprasegmental systems of Mandarin.

Smith’s definition of intelligibility also seems to be an attempt to fix the meaning of intelligibility to only countable items in an utterance. In other words, it is as if to say we can make intelligibility more “objective” and less “subjective” if we only count the phonemes or words that are clearly recognised by a NS listener. However, realistically the concept of “intelligibility” is really more subjective than objective especially if the NS listeners are not practitioners or researchers in the field of SLA. In this study, the NS markers are specifically recruited from the mainstream native speaking community. Although they are highly educated, they do not work in the field of language education. Consequently, their understanding of “intelligibility” in Mandarin is likely to be more subjective. Therefore, to prevent different interpretation of the concept of intelligibility influencing the marking process, markers were told to mark the conversations on an intelligibility level which indicates how natural the markers feel about a students’ overall oral performance compared to the native standard only. They were not asked to judge how correct or incorrect a students’ production of tone was as NSs who are inexperienced in teaching Mandarin to foreigners would certainly have different
interpretations of whether a tone is correct or incorrect and in what way they were correct or incorrect.

5.5 Quantitative Analysis

The statistical analyses were performed using the Analysis Toolpak in the Excel package and SPSS statistical package. The critical significance level was set at p<0.05 throughout the study. All oral production data from the 22 subjects were analysed. Statistical tests such as T-Test and Analysis of Variance (ANOVA) (Hatch & Lazaraton, 1991) were used to analyse the various characteristics of the voice ranges of all the subjects involved in this study.

The tonal productions of the 7 male subjects of the control and the 5 female subjects from the experimental group, 4 female subjects from experimental group and 5 female subjects from control group were analysed using Praat. On a frame-by-frame basis, the fundamental frequency was determined for each voiced frame of each subject’s test data. The frequency range for males was set from 80 Hz to 250 Hz and female range was from 150 Hz to 400 Hz. From the collection of F0 values for each subject, the mean, standard deviation, minimum, maximum and range were determined for that subject. In turn, the group averages were determined for those 5 variables. The tonal productions of the female subjects from both control and experimental groups were treated in the same manner.

T-test was also used to find whether there were significant differences in the ratings of the nine NSs of Mandarin. The consistency of the NS ratings of the students’ oral performance was also calculated using Cronbach Alpha reliability test in SPSS.
The quality of the conversation produced by both groups was also measured by Mean Length of Utterances (MLU) and creativity (measured by number of long sentences and use of language items outside the teaching materials).

Finally, errors made by both groups of students obtained through auditory analysis were described according to their tonal categories.

5.6 Qualitative analysis

Questionnaires administered at the end of the semester and one to one interviews with the researcher on the learning process were also used to elicit learning strategies that learners used in learning Mandarin and their opinions on improving the design of the subject. With the students in the control group, an end of semester oral test was conducted at the end of the first semester but the one to one interview was only conducted at the end of the second semester. Consequently, only four students’ interview data have been analysed. For the experimental group of students, both the questionnaire (see Appendix 8) and one to one interview were conducted at the end of the first semester in 2003-4 (see Appendix 4 for the guiding questions used in the interviews). From the experimental group, data was also collected from interviews from students whose oral productions were not analysed in the quantitative analyses. Due to the small number of interviews involved in both groups, analysis mainly followed the methodology of profiling the responses of students. The purpose of the interview data analysis was to triangulate with the data collected through quantitative means.

5.7 Summary

This chapter is concerned with the description of the empirical study undertaken to evaluate the efficacy of SEA in Mandarin learning. Details about the subjects in the study, methods of data
collection used in qualitative and quantitative analysis and instrumentation and research procedures are provided in this chapter.

In the next chapter, the quantitative results from both control and experimental groups will be reported.
Chapter 6 Quantitative results of the Study

6.1 Objectives

As has been stated in the Introduction of this thesis, a major aim of the study is to obtain information about learner characteristics on the acquisition of Mandarin pronunciation, focusing on Mandarin tones and prosody, during the initial stage of acquisition. The types of information sought include:

1. Sequence of mastery of the four tones, and hence the order of difficulty of those tones;
2. Common tone and prosodic error patterns among learners;
3. Likely cause(s) leading to the detected errors;
4. How and where interference from English intonation takes place and how significant its role;
5. Effectiveness of SEA on the teaching of Mandarin tones and prosody;
6. Effectiveness of specific computer enhanced learning programs in the learning process.

In this chapter, the results concerning items 1 to 5 in the above list are provided.

6.1.1 The acoustic experiment (Experiment one)

The first experiment consisted of an acoustic experiment conducted with 22 L2 students involved in the study. These 22 L2 students are composed of 5 male and 5 female subjects in the control (CG) and 8 male and 4 female subjects in the experimental group (EG). In previous acoustic studies of tones (Howie, 1974, 1976), fundamental frequency ($F_0$) has been considered the most important factor in tone perception. Therefore, in the acoustic experiment (Experiment one), the group average of L2 speakers’ mean $F_0$, minimum $F_0$ (i.e. the average starting $F_0$ of the group’s voice level), maximum $F_0$ (i.e. the average maximum level of $F_0$ for the group) and the average $F_0$ range for the L2 speakers are compared with those of NSs’. The aim of the
comparison is to seek evidence of any voice quality changes that might be attributed to SEA. The results obtained from the acoustic experiment (Experiment one) represent objective data obtained on any voice quality changes of L2 students.

6.1.2 The perceptual experiment (Experiment two)

In the perceptual experiment (Experiment two), the same speech productions analysed in Experiment one were given to nine female NSs of Mandarin. The researcher was not one of the nine judges. The NS judges were highly educated NSs of Mandarin from Beijing who were asked to listen to the speech production with no script and mark the “naturalness” of the speech from each student according to a scale of 1 to 9. They were not specifically asked to mark tones or intonation or prosody as any judgment of these aspects of the language required specialized knowledge of the Mandarin phonological system. Though highly educated, the NS judges who volunteered for Experiment two are unlikely to possess such specialized understanding of the Mandarin phonological system. “Naturalness” is defined as how close the utterances are to native speaker speech in terms of rhythm, tones, intonation, stress and discourse features.

The results obtained from the perceptual experiment (Experiment two) represent subjective judgment of L2 students’ performances from NSs of Mandarin. The judges did not know which groups (experimental or control) each of the students belonged to.

6.1.3 The auditory analysis

In order to explore in more detail the nature of L2 learner errors in Mandarin, an auditory analysis was also conducted. This analysis was done by the researcher (who is also a NS of Mandarin) and two other experienced teachers of Mandarin. The researcher has had 16 years of experience in teaching Mandarin to English speaking University students in Australia. The
other two teachers have been teaching Mandarin to English learners of Mandarin for a total of six years. Thus all three analysts possess a reasonable amount of specialized knowledge of Mandarin phonology. They separately marked, according to their judgment, the deviant tonal production of each syllable using Chao’s tone letter system (Chao, 1930). They were allowed to listen to each utterance as often as they wanted. A tone was considered mispronounced only when judged as inaccurate by at least 2 listeners. Based on the auditory analysis, an acoustic study was carried out to verify the mispronounced tones. The tonal productions of 4 subjects from the control and 5 subjects from the experimental group were analysed using Praat (Boersma & Weenink, 2003) to extract the overall mean F₀ of the speech samples. The starting-F₀ and the ending-F₀ of mispronounced tones, made by individual students and grouped by tonal category (except for the neutral tone 5 and Tone 3), were also analysed. The F₀ of the voices of the subjects and the mean F₀ values of Tones 1, 2 and 4 were computed respectively. From the detailed auditory analysis, it was possible to observe the order of difficulty of tones and the patterns of errors made by the students.

6.2 Results of the acoustic experiment (Experiment one)

6.2.1 Fundamental frequency (F₀) of the male subjects

6.2.1.1 The experimental group of male subjects

Table 6.1 The group averages of speakers’ mean F₀, mean F₀ SD, semitone, mean minimum F₀, mean maximum F₀ and mean F₀ range for 5 groups of male speakers

<table>
<thead>
<tr>
<th>Cohort</th>
<th>Mean F₀</th>
<th>SD</th>
<th>Semitone</th>
<th>Min</th>
<th>Max</th>
<th>Range</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>NSM</td>
<td>131.52</td>
<td>30.83</td>
<td>3.32</td>
<td>80.23</td>
<td>268.19</td>
<td>187.97</td>
<td>13</td>
</tr>
<tr>
<td>EG(Ch)</td>
<td>130.67</td>
<td>27.42</td>
<td>3.28</td>
<td>82.92</td>
<td>277.81</td>
<td>194.89</td>
<td>8</td>
</tr>
<tr>
<td>EG(En)</td>
<td>115.96</td>
<td>21.01</td>
<td>2.69</td>
<td>84.78</td>
<td>258.99</td>
<td>174.20</td>
<td>8</td>
</tr>
<tr>
<td>CG(Ch)</td>
<td>115.49</td>
<td>20.57</td>
<td>2.77</td>
<td>84.14</td>
<td>245.54</td>
<td>161.40</td>
<td>5</td>
</tr>
<tr>
<td>CG(En)</td>
<td>110.16</td>
<td>16.68</td>
<td>2.17</td>
<td>80.92</td>
<td>235.05</td>
<td>218.37</td>
<td>5</td>
</tr>
</tbody>
</table>

Key: NSM: native speaker male speakers speaking Mandarin
EG (Ch): male experimental group speakers speaking Mandarin
EG (En): male experimental group speakers speaking English
CG (Ch): male control group speakers speaking Mandarin
CG (En): male control speakers speaking English
From Table 6.1, the following results emerge:

1. All male English speaking subjects, in both experimental and control groups, had narrower pitch ranges in their native language, English than in Mandarin. The experimental and control groups have group average F0 ranges of 174.20 Hz and 218.37 Hz respectively.

2. In contrast, the group average F0 range of the Mandarin male subjects (NSM) speaking their native language (187.97 Hz) is greater than that of the experimental group speaking their native language (174.20 Hz) but smaller than the control speaking their native language (218.37 Hz).

3. For the experimental group of male students, their average group F0 range increased from 174.20 Hz to 194.89 Hz when switching from English to Mandarin. However, the experimental group had a smaller average group F0 range in their native language (174.20 Hz) than the NSs’ pitch range in Mandarin (187.97 Hz). This supports Chen’s finding that “When the four English-speaking subjects switched from their native language to Chinese, their pitch range increased….when utterance sentences were used” (Chen, 1974).

However, as the speech samples analysed in this study consisted of sentences rather than words, a more reliable measure is perhaps to look at the group average mean F0 (\( \bar{X} \)) for the group. An examination of the experimental group’s English and Mandarin group average mean F0 (\( \bar{X} \)) revealed that when switching from English to Mandarin, the \( \bar{X} \) increased by 14.71 Hz (130.67-115.96 Hz). Similarly, the group average F0 SD (group average standard deviation of the average mean F0 of the group) also widened when changing from English to Chinese. This is evident by the increase in average F0 SD of 6.41 Hz (27.42 Hz-21.01 Hz). The group average F0 SD is an acoustic measure measuring the distribution of average F0 in the group.
In order to ascertain whether the differences in the group mean $F_0$, group average $F_0$ range, and group average $F_0$ SD of Mandarin and English for the experimental group were statistically significant, t-tests were carried out for those variables. A test of equality of variances was also carried out for each analysis in order to make sure that the variances between these variables were comparable. For each analysis, this test was passed. Table 6.2 shows the difference in the group average mean $F_0$ of the male students in the experimental group speaking English and Mandarin was significantly different at $p<0.01$ level with $p=0.000$. The difference in the group average $F_0$ SD was nearly reaching significance at $p<0.05$ level with $p=0.060$. The difference in semitone which measures the liveliness of the conversation was significant at $p<0.05$ level with $p=0.046$ level.

Table 6.2 T-tests for group average mean $F_0$, average mean $F_0$ SD and semitone, minimum $F_0$, maximum $F_0$, $F_0$ range, of the male speakers in the experimental group, for Mandarin and English languages (in Hz)

<table>
<thead>
<tr>
<th>Languages</th>
<th>Mean $F_0$</th>
<th>Mean $F_0$ SD</th>
<th>Semitone</th>
<th>Minimum $F_0$</th>
<th>Maximum $F_0$</th>
<th>$F_0$ range</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mandarin</td>
<td>130.67</td>
<td>27.42</td>
<td>3.28</td>
<td>82.92</td>
<td>277.81</td>
<td>194.89</td>
</tr>
<tr>
<td>English</td>
<td>115.96</td>
<td>21.01</td>
<td>2.69</td>
<td>84.78</td>
<td>258.9888</td>
<td>174.20</td>
</tr>
</tbody>
</table>

**Significance Level (p)**
- **0.000**
- 0.060
- 0.046*
- 0.32
- 0.30
- 0.24

** indicates that the differences are statistically significant at $p=0.01$ level.

*: indicates that the differences are statistically significant at $p=0.05$ level.

Firstly, this means that, after training with SEA, the experimental group of learners spoke Mandarin at a higher level of pitch (measured by the group average mean $F_0$) than when they spoke English.

Secondly, not only did the experimental group manage to speak Mandarin at a higher average pitch within each person’s speech, they were able to stretch their pitch range. In other words, they managed to expand their average pitch (measured in group average mean $F_0$ SD: standard deviation of the fundamental frequency). This could mean that they were trying to accommodate for Mandarin tones.
Thirdly, although the pitch ranges of this group speaking Mandarin and English are not significantly different, the statistically significant difference in the group average mean F0 and near statistical significant differences of the width of the pitch range (group average mean F0 SD) and the statistical significant difference in semitone indicated that there were significant changes in this group of students’ voice quality when they speak Mandarin.

6.2.1.2 The control group male subjects

In contrast, according to Table 6.1, for the control group of male students, their group average pitch range decreased from 218.37 Hz to 161.40 Hz when they switched from English to Mandarin. Their group average pitch range in English (218.37 Hz) was also wider than the NSs’ pitch range in Mandarin (187.97 Hz).

In order to ascertain whether the differences in the group mean F0, group average F0 range, group average F0 SD, semitone, minimum F0 and maximum F0 of Mandarin and English for the male students in the control group were statistically significant, t-tests were carried out for those variables. A test of equality of variances was also carried out for each analysis. For each analysis, this test was passed. Table 6.3 shows the differences in all the variables in English and Mandarin were statistically insignificant. This means that when switching from English to Mandarin, the changes in these aspects of the voice quality of the control group male subjects were not significant. There might be a gender-based explanation for the control group of males’ poor ability to speak Mandarin. As speaking Mandarin requires the speakers to be more expressive as shown by the higher semitone required, Australian males might not know how to be expressive or might be reluctant to be too expressive to accommodate the tonal nature of Mandarin as speaking at a high pitch is considered “girlie” culturally. Consequently without undergoing a period of consistent, overt perception and production training in Mandarin (such
as what the experimental group went through), males students in the control group obviously found it hard to express Mandarin in a lively manner or with higher average $F_0$ as required by the characteristics of Mandarin.

Table 6.3 T-tests for group average mean $F_0$, average mean $F_0$ SD and semitone, minimum $F_0$, maximum $F_0$, $F_0$ range, of the male speakers in the control group, for Mandarin and English languages (in Hz)

<table>
<thead>
<tr>
<th>Languages</th>
<th>Mean $F_0$</th>
<th>Mean $F_0$ SD</th>
<th>Semitone</th>
<th>Minimum $F_0$</th>
<th>Maximum $F_0$</th>
<th>$F_0$ range</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mandarin</td>
<td>115.49</td>
<td>20.57</td>
<td>2.62</td>
<td>84.14</td>
<td>245.54</td>
<td>161.40</td>
</tr>
<tr>
<td>English</td>
<td>110.16</td>
<td>16.68</td>
<td>2.18</td>
<td>80.92</td>
<td>235.05</td>
<td>218.37</td>
</tr>
</tbody>
</table>

| Significance Level (p) | 0.34 | 0.42 | 0.33 | 0.17 | 0.76 | 0.16 |

6.2.1.3 Liveliness of discourse

Traumuller and Eriksson pointed out that the statistics of $F_0$-values are often not very well described by a normal distribution. If $F_0$ is scaled linearly (in Hz), there is, typically, some positive skewness (Traunmuller & Eriksson, 1995). Therefore, in order to compare the results from different groups in this study, the average mean $F_0$ SD for the group is also reported in semitones. Traunmuller and Eriksson further point out that the liveliness of a discourse can be observed by comparing the $F_0$-excursion of the utterances measured in semitones. In other words, the higher the SD of the average mean $F_0$ measured in semitones, the livelier is the discourse (Traunmuller & Eriksson, 1995). According to Table 6.1, all male subjects in both experimental and control students spoke Mandarin with a livelier discourse than when they spoke English. Subjects in the experimental group spoke Mandarin with a liveliness level (SD in semitone: 3.28 semitone) at almost the same level as the NS group (SD in semitone: 3.32 semitone). However, the subjects in the control group spoke Mandarin with a liveliness level
(SD in semitone: 2.77 semitone), which was lower than that of the NS group (SD in semitone: 3.32 semitone). Therefore, we can conclude that in order to speak Mandarin with intelligible prosody and tones, it is necessary for students to be more expressive in the way they speak.

### 6.2.1.4 Is there a difference between the experimental, control and NS groups’ behaviour in speaking Mandarin?

It can be hypothesized that if the male students in the experimental group of students spoke with expanded pitch range and a higher average mean $F_0$, the differences in the average mean $F_0$ and average mean $F_0$ SD (the width of the pitch range) should be comparable to that of the male NS group. An Analysis of Variance was thus carried out to discover the relationship between different groups of subjects. This was done by conducting a One-way ANOVA test. The use of one way ANOVA to make between-group comparison was recommended and approved by a qualified statistician.

Table 6.4 One-way ANOVA test comparing group average mean $F_0$, average mean $F_0$ SD and semitone, minimum $F_0$, maximum $F_0$, $F_0$ range, of the male speakers in all groups in Mandarin (in Hz)

<table>
<thead>
<tr>
<th>Groups</th>
<th>Mean $F_0$</th>
<th>Mean $F_0$ SD</th>
<th>Semitone</th>
<th>Minimum $F_0$</th>
<th>Maximum $F_0$</th>
<th>$F_0$ range</th>
</tr>
</thead>
<tbody>
<tr>
<td>Experimental</td>
<td>130.67</td>
<td>27.42</td>
<td>3.28</td>
<td>82.92</td>
<td>277.81</td>
<td>194.89</td>
</tr>
<tr>
<td>Control</td>
<td>115.49</td>
<td>20.57</td>
<td>2.77</td>
<td>84.14</td>
<td>245.54</td>
<td>161.40</td>
</tr>
<tr>
<td>Native Speaker</td>
<td>131.52</td>
<td>30.83</td>
<td>3.32</td>
<td>80.23</td>
<td>268.19</td>
<td>187.97</td>
</tr>
</tbody>
</table>

Significance Level (p) | 0.070 | 0.071 | 0.029* | 0.16 | 0.32 | 0.30 |

*: indicates that the differences are statistically significant at p=0.05 level.

Table 6.4 shows that there are differences between the three groups in terms of the group average mean $F_0$, the group average mean $F_0$ SD and the semitone with the differences nearly reaching statistical significance at p<0.05 level with p=0.070 and p=0.071 respectively. The differences in semitone were statistically different at p<0.05 level with p=0.029.

Post Hoc tests for the one way ANOVA between the three groups were also carried out. Table 6.5 shows that there is no difference between the average mean $F_0$ between the experimental
and the NS groups with p=1 (indicated in boldface). Between the control and NS groups, however, the difference is not highly significant although differences do exist (p=0.084) (indicated in boldface)

Table 6.5 Post Hoc tests for one way ANOVA for the dependent variable average mean F_0 for Mandarin between experimental, control and NS groups of male subjects.

<table>
<thead>
<tr>
<th>(I) GROUP</th>
<th>(J) GROUP</th>
<th>Mean Difference (I-J)</th>
<th>Std. Error</th>
<th>Sig.</th>
<th>Lower Bound</th>
<th>Upper Bound</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.00</td>
<td>2.00</td>
<td>15.1760</td>
<td>7.30206</td>
<td>.147</td>
<td>-3.6780</td>
<td>34.0300</td>
</tr>
<tr>
<td>3.00</td>
<td>1.00</td>
<td>-15.1760</td>
<td>7.30206</td>
<td>.147</td>
<td>-34.0300</td>
<td>14.2382</td>
</tr>
<tr>
<td>3.00</td>
<td>2.00</td>
<td>-15.7991</td>
<td>6.74036</td>
<td>.084</td>
<td>-33.2028</td>
<td>1.6047</td>
</tr>
<tr>
<td>2.00</td>
<td>1.00</td>
<td>.6231</td>
<td>5.75568</td>
<td>1.000</td>
<td>-14.2382</td>
<td>15.4843</td>
</tr>
</tbody>
</table>

Key: Group 1: the experimental group of male subjects.
Group 2: the control group male subjects.
Group 3: the NS group of male subjects.

Table 6.6 Post Hoc tests for one way ANOVA for the dependent variable average mean F_0 SD for Mandarin between experimental, control and NS groups of male subjects.

<table>
<thead>
<tr>
<th>(I) GROUP</th>
<th>(J) GROUP</th>
<th>Mean Difference (I-J)</th>
<th>Std. Error</th>
<th>Sig.</th>
<th>Lower Bound</th>
<th>Upper Bound</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.00</td>
<td>2.00</td>
<td>6.8480</td>
<td>4.56335</td>
<td>.441</td>
<td>-4.9347</td>
<td>18.6307</td>
</tr>
<tr>
<td>3.00</td>
<td>1.00</td>
<td>-6.8480</td>
<td>4.56335</td>
<td>.441</td>
<td>-18.6307</td>
<td>5.8782</td>
</tr>
<tr>
<td>2.00</td>
<td>3.00</td>
<td>-10.2572</td>
<td>4.21233</td>
<td>.069</td>
<td>-21.1335</td>
<td>.6191</td>
</tr>
<tr>
<td>3.00</td>
<td>2.00</td>
<td>3.4092</td>
<td>3.59696</td>
<td>1.000</td>
<td>-5.8782</td>
<td>12.6966</td>
</tr>
</tbody>
</table>

Key: Group 1: the experimental group of male subjects.
Group 2: the control group male subjects.
Group 3: the NS group of male subjects.
Table 6.7 Post Hoc tests for one way ANOVA for the dependent variable semitone for Mandarin between experimental, control and NS groups of male students.

<table>
<thead>
<tr>
<th>(I) GROUP</th>
<th>(J) GROUP</th>
<th>Mean Difference (I-J)</th>
<th>Std. Error</th>
<th>Sig.</th>
<th>95% Confidence Interval</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.00</td>
<td>2.00</td>
<td>.6490</td>
<td>.44075</td>
<td>.463</td>
<td>-.4890 - 1.7870</td>
</tr>
<tr>
<td>3.00</td>
<td>3.00</td>
<td>-.5012</td>
<td>.34741</td>
<td>.488</td>
<td>-1.3982 - .3959</td>
</tr>
<tr>
<td>2.00</td>
<td>1.00</td>
<td>-.6490</td>
<td>.44075</td>
<td>.463</td>
<td>-1.7870 - .4890</td>
</tr>
<tr>
<td>3.00</td>
<td>2.00</td>
<td>-1.15(*)</td>
<td>.40685</td>
<td>.029*</td>
<td>-2.2006 - .0997</td>
</tr>
<tr>
<td>3.00</td>
<td>1.00</td>
<td>.5012</td>
<td>.34741</td>
<td>.488</td>
<td>-.3959 - 1.3982</td>
</tr>
<tr>
<td>2.00</td>
<td>2.00</td>
<td>1.1502(*)</td>
<td>.40685</td>
<td>.029*</td>
<td>.0997 - 2.2006</td>
</tr>
</tbody>
</table>

*: indicates that the differences are statistically significant at p=0.05 level.

Key: Group 1: the experimental group of male subjects.
Group 2: the control group male subjects.
Group 3: the NS group of male subjects.

Similarly, in Table 6.6 and Table 6.7, the differences in average mean F0 SD and semitone were nearly significant or significant between the control and NS groups (p=0.069 and 0.029 respectively) and not significant between the experimental and NS groups (p=1 and 0.488 respectively). These statistical tests offer some evidence that the voice quality of male subjects in the experimental group is similar to that of NS males when speaking Mandarin. However, the voice quality of the male subjects in the control is significantly different from the experimental and NS groups.
6.2.2 Fundamental Frequency (F\(_0\)) of the female subjects

6.2.2.1 The experimental group of female subjects

Table 6.8 The group averages of speakers’ mean F\(_0\), minimum F\(_0\), maximum F\(_0\) and F\(_0\) range for 5 groups of female speakers

<table>
<thead>
<tr>
<th>Groups</th>
<th>Mean F(_0)</th>
<th>Mean F(_0) SD</th>
<th>Semitone</th>
<th>Minimum F(_0)</th>
<th>Maximum F(_0)</th>
<th>F(_0) range</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>NSF(Ch)</td>
<td>207.09</td>
<td>41.92</td>
<td>3.32</td>
<td>148.39</td>
<td>351.19</td>
<td>202.80</td>
<td>6</td>
</tr>
<tr>
<td>EGF(Ch)</td>
<td>212.81</td>
<td>40.03</td>
<td>2.97</td>
<td>150.94</td>
<td>362.96</td>
<td>236.02</td>
<td>4</td>
</tr>
<tr>
<td>EGF(En)</td>
<td>184.62</td>
<td>30.50</td>
<td>2.48</td>
<td>149.96</td>
<td>352.83</td>
<td>202.87</td>
<td>4</td>
</tr>
<tr>
<td>CGF(Ch)</td>
<td>208.06</td>
<td>35.58</td>
<td>2.77</td>
<td>152.00</td>
<td>379.77</td>
<td>227.77</td>
<td>5</td>
</tr>
<tr>
<td>CGF(En)</td>
<td>181.49</td>
<td>26.26</td>
<td>2.17</td>
<td>151.48</td>
<td>358.34</td>
<td>206.86</td>
<td>5</td>
</tr>
</tbody>
</table>

Key: NSF: NS female speakers speaking Mandarin
EGF(Ch): female experimental group speakers speaking Mandarin
EGF(En): female experimental group speakers speaking English
CGF(Ch): female control speakers speaking Mandarin
CGF(En): female control speakers speaking English

From Table 6.8

All the female English speaking subjects, in both experimental and control group, had narrow pitch ranges in their native language, English, with a group average mean F\(_0\) range of 202.87 Hz for the experimental group and 206.86 Hz for the control group. The group average F\(_0\) range of the Mandarin female subjects speaking their native language is similar to both experimental and control groups with an average F\(_0\) range of 202.80 Hz.

For the experimental group of female students, their average mean F\(_0\) increased from 184.62 Hz to 212.81 Hz when they switched from English to Mandarin. However, the experimental group had a smaller average group F\(_0\) in their native language (184.62 Hz) than the NSs’ pitch range in Mandarin (207.09 Hz). This again supports Chen’s finding (Chen, 1974).

An examination of the female subjects in the experimental group’s English and Mandarin group average mean F\(_0\) (\(\overline{X}\)) revealed that when switching from English to Mandarin, the \(\overline{X}\) increased 28.19 Hz (184.62 Hz-212.81 Hz). Similarly, the group average F\(_0\) SD (the average mean F\(_0\) SD)
also widened when changing from English to Chinese. This is evident by the increase in average F₀ SD of 9.53 Hz (40.03 Hz-30.50 Hz).

In order to ascertain whether the differences in the group mean F₀, group average F₀ range, and group average F₀ SD of Mandarin and English for the experimental group of female speakers were statistically significant, T-tests were carried out. A test of equality of variances was also carried out for each analysis of variance. In each analysis, this test was passed. Table 6.9 shows the difference in the group average mean F₀ of the experimental group speaking English and Mandarin was not significantly different at p<0.05 level with p=0.098. In all other aspects, the differences were also not significantly different.

<table>
<thead>
<tr>
<th>Languages</th>
<th>Mean F₀</th>
<th>Mean F₀ SD</th>
<th>Semitone</th>
<th>Minimum F₀</th>
<th>Maximum F₀</th>
<th>F₀ range</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mandarin</td>
<td>212.81</td>
<td>40.03</td>
<td>2.9725</td>
<td>150.94</td>
<td>386.9625</td>
<td>236.023</td>
</tr>
<tr>
<td>English</td>
<td>184.62</td>
<td>30.50</td>
<td>2.48</td>
<td>149.96</td>
<td>353.83</td>
<td>202.87</td>
</tr>
<tr>
<td>Significance Level (p)</td>
<td>0.098</td>
<td>0.489</td>
<td>0.566</td>
<td>0.470</td>
<td>0.347</td>
<td>0.369</td>
</tr>
</tbody>
</table>

There is no significant difference in the experimental group of females’ speech when switching from English to Mandarin although some differences can be observed with regards to the group average mean F₀. An explanation could be that when speaking Mandarin, female subjects in the experimental group spoke at a higher mean F₀ than when they spoke English but the differences are not highly significant with p=0.098.
6.2.2.2 The control group female subjects

Table 6.10 T-tests for group average mean F₀, average mean F₀ SD and semitone, minimum F₀, maximum F₀, F₀ range, of the female speakers in the control group, for Mandarin and English languages (in Hz)

<table>
<thead>
<tr>
<th>Languages</th>
<th>Mean F₀</th>
<th>Mean F₀ SD</th>
<th>Semitone</th>
<th>Minimum F₀</th>
<th>Maximum F₀</th>
<th>F₀ range</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mandarin</td>
<td>208.056</td>
<td>35.580</td>
<td>2.77</td>
<td>152.00</td>
<td>379.769</td>
<td>227.770</td>
</tr>
<tr>
<td>English</td>
<td>181.49</td>
<td>26.26</td>
<td>2.17</td>
<td>151.48</td>
<td>358.34</td>
<td>208.06</td>
</tr>
</tbody>
</table>

Significance level (p)

<table>
<thead>
<tr>
<th>Languages</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mandarin</td>
<td>0.032*</td>
</tr>
<tr>
<td>English</td>
<td>0.122</td>
</tr>
</tbody>
</table>

*: indicates that the differences are statistically significant at p=0.05 level.

There is no significance difference in the speech from the female subjects in the control group except in terms of their group average mean F₀ (Mean F₀: p=0.032) when switching from English. This means that when speaking Mandarin, female subjects in the control group generally spoke at a significantly higher mean F₀ than when they spoke English. At first glance, it seems that just by learning Mandarin, female speakers in the control group seem to have significantly altered the way they speak. However, the data contained in Table 6.10 shows only the group average mean F₀ as significantly different. No significant differences have been observed for the group Mean F₀ SD, minimum F₀, maximum F₀ or range of F₀. This differs from the experimental male group’s data shown in Table 6.2. In Table 6.2, highly significant and nearly significant differences have been observed in the group mean F₀, Mean F₀ SD and Semitone of the experimental male group speakers with p=0; 0.060 and 0.046 respectively.

The fact that there are no significant differences between the group Mean F₀ SD and semitone for the control group female speakers suggests that female speakers might have over-generalized by speaking Mandarin at a higher average mean F₀ at all times. They did not know
when to speak at a higher frequency. Even in Mandarin, it is not required that one speaks Mandarin at a higher frequency, i.e. with a higher pitch 100% at a time.

6.2.2.3 Is there a difference between the female subjects of the experimental, control and NS groups’ behaviour in speaking Mandarin?

It can be hypothesized that if there are significant differences in the way experimental group of female students spoke Mandarin and English, the differences in the average mean $F_0$ and average mean $F_0$ SD (the width of the pitch range) in Mandarin should be comparable to that of the NS group. An Analysis of Variance was thus carried out to discover the relationship between different groups of subjects.

A comparison between the experimental, control and NS groups of female speakers in speaking Mandarin was also carried out. This was done by conducting a one-way ANOVA test.

Table 6.11 One-way ANOVA test comparing group average mean $F_0$, average mean $F_0$ SD and semitone, minimum $F_0$, maximum $F_0$, $F_0$ range, of the female speakers in all groups in Mandarin (in Hz)

<table>
<thead>
<tr>
<th>Groups</th>
<th>Mean $F_0$</th>
<th>Mean $F_0$ SD</th>
<th>Semitone</th>
<th>Minimum $F_0$</th>
<th>Maximum $F_0$</th>
<th>$F_0$ range</th>
</tr>
</thead>
<tbody>
<tr>
<td>Experimental</td>
<td>212.81</td>
<td>40.03</td>
<td>2.97</td>
<td>150.94</td>
<td>386.96</td>
<td>236.02</td>
</tr>
<tr>
<td>Control</td>
<td>208.06</td>
<td>35.58</td>
<td>2.77</td>
<td>152.00</td>
<td>379.77</td>
<td>227.77</td>
</tr>
<tr>
<td>NS Speaker</td>
<td>207.09</td>
<td>41.92</td>
<td>3.32</td>
<td>148.39</td>
<td>351.19</td>
<td>202.80</td>
</tr>
</tbody>
</table>

| Significance Level (p) | 0.800  | 0.62  | 0.43  | 0.036* | 0.029* | 0.062 |

*: indicates that the differences are statistically significant at p=0.05 level.

Comparing the group mean $F_0$, group average $F_0$ range, and group average $F_0$ SD of experimental, control groups with the NS group of female subjects, it is found that there were significant differences in the minimum and maximum $F_0$ when speaking Mandarin between the groups.
Post Hoc tests for the one way ANOVA between the three groups of female subjects show that minimum $F_0$ between the experimental and the NS groups is not statistically different with $p=0.243$ (indicated in boldface). Between the control and NS groups, however, the difference is highly significant differences exist ($p=0.042$) (indicated in boldface). See Table 6.12.

Table 6.12 Post Hoc tests for one way ANOVA for the dependent variable minimum $F_0$ for Mandarin between experimental, control and NS groups of female subjects.

<table>
<thead>
<tr>
<th>(I) GROUP</th>
<th>(J) GROUP</th>
<th>Mean difference (I-J)</th>
<th>Std. Error</th>
<th>Sig.</th>
<th>95% Confidence Interval</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Lower bound</td>
</tr>
<tr>
<td>1</td>
<td>2</td>
<td>-1.0620</td>
<td>1.39169</td>
<td>1.000</td>
<td>-4.9302</td>
</tr>
<tr>
<td>1</td>
<td>3</td>
<td>2.5517</td>
<td>1.33915</td>
<td>.243</td>
<td>-1.1705</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>1.0620</td>
<td>1.39169</td>
<td>1.000</td>
<td>-2.8062</td>
</tr>
<tr>
<td>2</td>
<td>3</td>
<td>3.6137(*)</td>
<td>1.25624</td>
<td>.042</td>
<td>.1220</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>-2.5517</td>
<td>1.33915</td>
<td>.243</td>
<td>-6.2738</td>
</tr>
<tr>
<td>3</td>
<td>2</td>
<td>-3.6137(*)</td>
<td>1.25624</td>
<td>.042</td>
<td>-7.1053</td>
</tr>
</tbody>
</table>

*: indicates that the differences are statistically significant at $p=0.05$ level.

Key:
Group 1: the experimental group of male subjects.
Group 2: the control group male subjects.
Group 3: the NS group of male subjects.

Table 6.13 shows that maximum $F_0$ between the experimental and the NS groups is statistically different with $p=0.047$ (indicated in boldface). Between the control and NS groups, however, the difference is not significant ($p=0.102$) (indicated in boldface).

The information in the Post Hoc tests suggests that the experimental group subjects reached a higher maximum $F_0$ when speaking Mandarin when compared to both the control and NS groups of female subjects (see Table 6.11 and Table 6.13). Furthermore, for both the experimental and control groups, their group average minimum $F_0$s are significantly higher than the NS group’s (see Table 6.11 and Table 6.13).
Table 6.13 Post Hoc tests for one way ANOVA for the dependent variable maximum F0 for Mandarin between experimental, control and NS groups of female subjects

<table>
<thead>
<tr>
<th>(I) GROUP</th>
<th>(J) GROUP</th>
<th>Mean Difference (I-J)</th>
<th>Std. Error</th>
<th>Sig.</th>
<th>95% Confidence Interval Lower Bound</th>
<th>95% Confidence Interval Upper Bound</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>7.1945</td>
<td>13.23378</td>
<td>1.000</td>
<td>-29.5884</td>
<td>43.9774</td>
</tr>
<tr>
<td>1</td>
<td>3</td>
<td>35.7742(*)</td>
<td>12.73421</td>
<td>.047</td>
<td>-.3798</td>
<td>71.1686</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>-7.1945</td>
<td>13.23378</td>
<td>1.000</td>
<td>-43.9774</td>
<td>29.5884</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>-35.7742(*)</td>
<td>12.73421</td>
<td>.047</td>
<td>-71.1686</td>
<td>-.3798</td>
</tr>
<tr>
<td>3</td>
<td>2</td>
<td>-28.5797</td>
<td>11.94575</td>
<td>.102</td>
<td>-61.7825</td>
<td>4.6232</td>
</tr>
</tbody>
</table>

*: indicates that the differences are statistically significant at p=0.05 level.

Key:
Group 1: experimental group of female subjects.
Group 2: control group female subjects.
Group 3: NS group of female subjects.

This could be explained by the possibility that female learners of Mandarin tended to over-generalize during the process of learning Mandarin. In other words, they realized that in order to incorporate tones, they had to speak in a higher voice but at the very beginning of learning, learners might not be clear when a higher pitch was necessary. So a general rule of speaking everything at a higher level was adopted during speaking. The fact that their minimum F0 is higher than that of the NSs is also expected as Tone 4 needs to be produced lower than what English female speakers can produce when speaking Mandarin.

6.2.3 The effects of Somatically Enhanced Approach on the Fundamental Frequency (F0) of the subjects

In section 6.2.1, nearly significant differences between the group average mean F0 and group average mean F0 SDs of all groups of male students are described. The most likely explanation for this could be that the experimental group managed to significantly widen their pitch range when speaking Mandarin because they learned Mandarin using gestures and movement in SEA. The similarity in the average mean F0 between the experimental group and the NS group means...
that students from the experimental group significantly widened their pitch range when speaking Mandarin. This is evidenced by the comparison of SD of the average mean F_0 of their English and Mandarin speech production. In Table 6.2, the difference in the average mean F_0 SD for Mandarin and English for the experimental group was significant at p<0.01 level. When speaking English, the average mean F_0 SD for the experimental group is 21.01 Hz. When speaking Mandarin, the group average mean F_0 SD for the male subjects in the experimental group is 27.42 Hz. This represented an increase of:

\[
\frac{(27.42-21.01)}{21.01} \times 100\% = 30.51\%.
\]

However, the increase in the experimental group’s average mean F_0 SD in Mandarin was not wide enough. Using the group average mean F_0 SD of NSs as a guide (30.83 Hz), further expansion of the experimental group’s average mean F_0 SD is required.

The statistically significant differences between the male subjects in the control and the NS group suggest that the average pitch range and the width of the pitch range of the students in the control group did not undergo significant expansion after learning through a communicative teaching approach for Mandarin. Further evidence comes from the comparison of the group average F_0 SD of the control group’s English and Mandarin speech production. When speaking English, the group average mean F_0 SD for the control group is 16.68 Hz. When speaking Mandarin, the group average of mean F_0 SD is 20.57 Hz. This amounts to an increase of:

\[
\frac{(20.57 \text{ Hz}-16.68 \text{ Hz})}{16.68} \times 100\% = 23.32\%.
\]

The smaller increase of the control group’s average mean F_0 SD of the male subjects (23.32%) tends to suggest that even though male L2 students in the control group were taught through a communicative approach, insufficient attention was paid to the prosody of Mandarin and tones.
They were only provided with the usual reminder from the teacher that tones are important. Therefore, in the case of the control group students, the smaller increase in the width of the pitch range could have been caused by students’ lack of confidence in knowing how to deal with the tonal and prosodic aspects of spoken Mandarin. Interview data (in Chapter 7) with this group of students supports this explanation. It revealed that though the importance of tones was recognized, it was only recognized on an abstract level. In other words, students only paid attention to tones when they were being reminded by the teacher in class. Once the teacher was absent, students tended to go back to the strategies they felt secure with. In other words, they went back to using their own romanization and characters to learn. When it came to speaking Mandarin, students found it hard to incorporate tones in their speech under examination conditions or during their daily practice of Mandarin. Without the diacritics, they felt lost as the only visual reminders of tones had been taken away. For further discussion, please see Chapter 7.

For the female subjects, though subjects in the experimental and control groups both spoke Mandarin at a lower minimum $F_0$ and higher maximum $F_0$, the differences between group average mean $F_0$, the group average mean $F_0$ SD, and average $F_0$ range of the two groups are not statistically different. This seems to suggest that in general, females start higher when speaking Mandarin but there is no expansion of the pitch range as indicated by the average mean $F_0$ SD in order to accommodate tones. However, although statistically insignificant, the average mean $F_0$, average mean $F_0$ SD, and $F_0$ range are higher in Mandarin than in English for both groups of students.

The fact that both female and male subjects in the experimental group spoke Mandarin at a higher average mean $F_0$ (i.e. pitch) could possibly be attributed to the effect of SEA. Similarly,
the nearly significant differences in the average mean F0 SD and semitone when speaking Mandarin and English for the male subjects in the experimental group provide further evidence that the use of SEA had been especially beneficial for the experimental group of male students.

Further evidence of the changes in the voice quality of the experimental subjects when speaking Mandarin comes from a comparison between the experimental, control and NS groups of male subjects. Such comparison of the male subjects reveals that when speaking Mandarin, the experimental group of male subjects’ voice quality was more like the NSs’. In contrast, the control group male subjects’ voice quality was very different from that of NSs. This evidence further lends support to the probability of SEA having an effect on the way subjects spoke Mandarin.

The findings of this study correlated to a certain extent with Chen’s findings (Chen, 1974). Chen’s study reported that when the four English-speaking subjects switched from English to Mandarin, their pitch range for utterance sentences increased by 62.7% for sentences. In this study, increases in pitch range occurred for both experimental and control groups, however not to the same extent.

It must be noted that there are significant differences between Chen’s study and this study:

1. Firstly, it must be born in mind the English-speaking subjects involved in this study were not fluent speakers of Mandarin (these students had only had 30 hours of face-to-face contact in studying Mandarin). Therefore, it was most unlikely that they have developed new awareness in the L2 yet. Consequently, measuring the speech range of these students using only the maximum and minimum frequency values of the speech samples would not be reliable as some students could occasionally produce high-pitched sounds. A more reliable measure of speech
range would be the average pitch (group average mean F₀) of the students. The three measures of changes in voice quality adopted in this thesis are the group average mean F₀, group average F₀ SD and semitone of the group average F₀ SD.

2. In Chen’s study (Chen, 1974), the English subjects all spoke Mandarin with no accent. Therefore, they can be called bilingual speakers of Mandarin and English. Consequently, it can be expected that their awareness of Mandarin is already stable. The measurement of the pitch range used in his study was done by calculating the difference in means of the maximum and minimum frequency values of the speech sounds in the speech samples rather than the difference in the average mean F₀ of the subjects involved. Secondly, the sample size in Chen’s study consisted of only four English speakers of Mandarin while in this study the number of subjects is larger (n=22).

3. Thirdly, in Chen’s study, both male and female data were measured together. In this study, 9 female and 13 male subjects’ data were measured separately in the analyses. Since female speakers are anatomically different from males, they usually have a higher average mean F₀ than male speakers. In acoustic studies of voice quality, female data and male data are usually analysed separately. In this study, comparisons were made by separating female data from the male data.

A closer examination of the data provided by Chen’s study showed that none of the differences exist in average mean F₀, pitch range (maximum F₀-minimum F₀) for both females and males. Only the average mean F₀ SD in Mandarin and English for the male subjects differed significantly (p=0.04) at p<0.05 level (see Table 6.14). The significant difference in the average mean F₀ SD in Mandarin for male bilingual speakers further supports the findings of this study that (1) the difference in the group average mean F₀ SD in English and Mandarin can be used as
an indicator of pitch change used to accommodate the prosody of Mandarin and (2) male English learners of Mandarin are the group that most likely needs to change their voice quality consistently and unconsciously in order to speak native level Mandarin; (3) the fact that male students taught by SEA have already changed their pitch range consistently after only 30 hours of contact constitute evidence for the efficacy of SEA.

The research shows that for English speakers who have acquired a near NS of Mandarin, the only measurable change is likely to be the pitch range (measured by the average mean F₀ SD) in Mandarin when compared to English. In other words, the tonal nature of Mandarin requires NNSs of Mandarin to develop a wider pitch range to accommodate the tones and other prosodic aspects of Mandarin.

Table 6.14 T-tests for group average mean F₀, average mean F₀ SD and minimum F₀, maximum F₀, F₀ range, of the male speakers in Chen’s study (in Hz) (n=2)

<table>
<thead>
<tr>
<th>Languages</th>
<th>Mean F₀</th>
<th>Mean F₀ SD</th>
<th>F₀ range</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mandarin</td>
<td>109.28</td>
<td>19.9200</td>
<td>61.12</td>
</tr>
<tr>
<td>English</td>
<td>109.85</td>
<td>15.255</td>
<td>85.3250</td>
</tr>
<tr>
<td>Significance Level (p)</td>
<td>0.871</td>
<td>0.04*</td>
<td>0.665</td>
</tr>
</tbody>
</table>

*: indicates that the differences are statistically significant at p=0.05 level.

Table 6.15 T-tests for group average mean F₀, average mean F₀ SD and minimum F₀, maximum F₀, F₀ range, of the female speakers in Chen’s study (in Hz) (n=2)

<table>
<thead>
<tr>
<th>Languages</th>
<th>Mean F₀</th>
<th>Mean F₀ SD</th>
<th>F₀ range</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mandarin</td>
<td>227.29</td>
<td>16.75</td>
<td>123.27</td>
</tr>
<tr>
<td>English</td>
<td>220.46</td>
<td>30.855</td>
<td>.67</td>
</tr>
<tr>
<td>Significance Level (p)</td>
<td>.654</td>
<td>.169</td>
<td>.167</td>
</tr>
</tbody>
</table>

The findings of this study supports Chen’s conclusion that if an Australian English speaking person, or possibly any other non-tonal language speaker wishes to learn to speak Mandarin successfully, there is a need to widen his/her normal pitch range. For Australian speakers’ of English, particularly males, because of the relaxed way of speaking Australian English normally adopted by these speakers, the need to widen their normal pitch range, to get used to speaking at
a higher pitch level and to get use to tensing up their bodies to produce high-pitched sounds in speaking Mandarin is even more necessary.

6.3 Results of the perceptual experiment (Experiment two)

![Listeners' ratings for each subject's Mandarin oral performance in Control and Experimental groups](image)

**Figure 6.1** NS judges’ ratings on each subject’s Mandarin oral performance in the control and experimental groups

**Table 6.16** T-test analysis for dependent sample s illustrating the relation between experimental and control groups and the dependent sample (mean perceptual rating scores given by NS judges).

<table>
<thead>
<tr>
<th>Groups of students</th>
<th>N</th>
<th>MEAN</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>control group</td>
<td>10</td>
<td>4.51</td>
<td>0.73</td>
</tr>
<tr>
<td>experimental group</td>
<td>12</td>
<td>5.33</td>
<td>0.82</td>
</tr>
</tbody>
</table>

The difference means of the perceptual rating scores given by listeners was significant at p<0.05 level (p=0.02)

Table 6.16 presents the mean perceptual rating scores given by all the markers. The researcher was not one of the markers in the Experiment. The researcher made sure the recording conditions, in terms of location and the microphone used, were the same.
Some students (especially the older students) tended to be hesitant in their oral performance. The hesitancy manifested in the unnecessary introduction of “ums and ahs”, repetition of speech items and unnatural pausing. These characteristics of non-native speech clearly affected markers’ ratings. It is then no surprise that weak students tended to make many non-native pauses in their speech thus attracting a lower perceptual rating from the markers.

To evaluate the reliability of the data obtained, it is necessary to assess the consistency of the perceptual ratings from the markers. An inter-rater coefficient was calculated. A Cronbach Alpha reliability test using SPSS was conducted and an inter-rater reliability rating obtained was 0.92. This indicates that the agreement between the 9 markers is very high. A high inter-rater reliability rating of 0.92 means that though there might have been differences in the way different judges judged the L2 students’ speech production, they were in agreement that the experimental group of L2 students’ performed better than the control group students in terms of the “naturalness” of their Mandarin speech.

6.4 Error analysis of the Tonal errors
6.4.1 Summary: control group

The 11 students made 326 errors out of a total of 1507 tokens. One extra female student’s data was included in this analysis but not in the acoustic analysis because her English voice data was not available. This amounts to an error rate of 22% from the oral input from each student in the control group. The oral input was not controlled as it would have been if they had read a controlled reading passage, the distribution of different tones within each student’s sample is varied. Table 6.17 and Figure 6.2 show the distribution of each tone in the samples for each student in the control group. Table 6.18 and Figure 6.3 show the error distribution of the control group.
Table 6.17 Distribution of different tones and tone errors in control group’s oral production.

<table>
<thead>
<tr>
<th></th>
<th>S1</th>
<th>S2</th>
<th>S3</th>
<th>S4</th>
<th>S5</th>
<th>S6</th>
<th>S7</th>
<th>S8</th>
<th>S9</th>
<th>S10</th>
<th>S11</th>
<th>Total</th>
<th>% tone</th>
<th>% Tone error</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tone</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>14</td>
<td>18</td>
<td>43</td>
<td>14</td>
<td>36</td>
<td>23</td>
<td>22</td>
<td>15</td>
<td>22</td>
<td>93</td>
<td>0</td>
<td>390</td>
<td>25%</td>
<td>26</td>
</tr>
<tr>
<td>2</td>
<td>18</td>
<td>17</td>
<td>30</td>
<td>6</td>
<td>26</td>
<td>13</td>
<td>13</td>
<td>11</td>
<td>30</td>
<td>28</td>
<td>0</td>
<td>218</td>
<td>15%</td>
<td>7</td>
</tr>
<tr>
<td>3</td>
<td>33</td>
<td>24</td>
<td>19</td>
<td>22</td>
<td>33</td>
<td>45</td>
<td>20</td>
<td>22</td>
<td>75</td>
<td>70</td>
<td>0</td>
<td>393</td>
<td>26%</td>
<td>30</td>
</tr>
<tr>
<td>4</td>
<td>31</td>
<td>28</td>
<td>63</td>
<td>26</td>
<td>34</td>
<td>39</td>
<td>29</td>
<td>33</td>
<td>34</td>
<td>96</td>
<td>93</td>
<td>506</td>
<td>34%</td>
<td>37</td>
</tr>
<tr>
<td>Total no. of tones and errors of the Control group</td>
<td>96</td>
<td>87</td>
<td>155</td>
<td>68</td>
<td>129</td>
<td>118</td>
<td>109</td>
<td>81</td>
<td>89</td>
<td>294</td>
<td>281</td>
<td>1507</td>
<td>100</td>
<td>22</td>
</tr>
</tbody>
</table>

Figure 6.2 Distribution of different tones and tone errors of control group by each subject.
Table 6.18 Error distribution per tone for each subject in the control group

<table>
<thead>
<tr>
<th>Subjects</th>
<th>Total Tones</th>
<th>Total no. of errors</th>
<th>Total no. of Errors (%)</th>
<th>Tone 1 errors (%)</th>
<th>Tone 2 errors (%)</th>
<th>Tone 3 errors (%)</th>
<th>Tone 4 errors (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>S1</td>
<td>96.00</td>
<td>23.00</td>
<td>23.96</td>
<td>4.35</td>
<td>13.04</td>
<td>56.52</td>
<td>26.09</td>
</tr>
<tr>
<td>S2</td>
<td>87.00</td>
<td>34.00</td>
<td>39.08</td>
<td>29.41</td>
<td>5.88</td>
<td>35.29</td>
<td>29.41</td>
</tr>
<tr>
<td>S3</td>
<td>155.00</td>
<td>23.00</td>
<td>14.84</td>
<td>40.43</td>
<td>4.35</td>
<td>21.74</td>
<td>43.48</td>
</tr>
<tr>
<td>S4</td>
<td>88.00</td>
<td>20.00</td>
<td>29.41</td>
<td>15.00</td>
<td>15.00</td>
<td>35.00</td>
<td>35.00</td>
</tr>
<tr>
<td>S5</td>
<td>129.00</td>
<td>15.00</td>
<td>11.63</td>
<td>20.00</td>
<td>6.67</td>
<td>26.67</td>
<td>46.67</td>
</tr>
<tr>
<td>S6</td>
<td>118.00</td>
<td>22.00</td>
<td>18.84</td>
<td>27.27</td>
<td>0.00</td>
<td>54.55</td>
<td>18.18</td>
</tr>
<tr>
<td>S7</td>
<td>109.00</td>
<td>34.00</td>
<td>31.19</td>
<td>17.65</td>
<td>5.88</td>
<td>50.00</td>
<td>26.47</td>
</tr>
<tr>
<td>S8</td>
<td>81.00</td>
<td>20.00</td>
<td>24.69</td>
<td>25.00</td>
<td>0.00</td>
<td>25.00</td>
<td>50.00</td>
</tr>
<tr>
<td>S9</td>
<td>89.00</td>
<td>85.00</td>
<td>95.51</td>
<td>40.00</td>
<td>7.06</td>
<td>31.58</td>
<td>42.11</td>
</tr>
<tr>
<td>S10</td>
<td>294.00</td>
<td>19.00</td>
<td>9.46</td>
<td>15.79</td>
<td>10.53</td>
<td>31.58</td>
<td>42.11</td>
</tr>
<tr>
<td>S11</td>
<td>281.00</td>
<td>31.00</td>
<td>11.03</td>
<td>25.81</td>
<td>9.68</td>
<td>16.13</td>
<td>48.39</td>
</tr>
<tr>
<td>total</td>
<td>1507</td>
<td>326</td>
<td>22</td>
<td>26.38</td>
<td>7.06</td>
<td>30.06</td>
<td>36.50</td>
</tr>
</tbody>
</table>

Figure 6.3 Error distribution per tone for the control group

Table 6.18 shows that a total of 1507 tones were produced by the 11 students in the control group. 26.38% of these tones were Tone 1 errors, 7.06% were Tone 2 errors, 30.06% were Tone 3 errors and 36.50% were Tone 4 errors. Out of the total number of tones, 326 tonal errors that occupy 22% of the total tones produced were made by students. Figure 6.2 shows the distribution of tones and tonal errors made by each subject in the control group. Figure 6.3 shows the distribution of tonal errors made by each subject in the control group.
The number of errors made by the subjects in the control group varied a great deal. Table 6.18 shows the distribution of the different tones and errors made by each subject in the control group. Only subject 9 (S9) made a large number of errors which constituted 95.51% of the language produced. On the other hand, subject 10 (S10), for example produced 294 syllables but only around 19 (6.46%) were tonal errors.

### 6.4.2 Summary: experimental group

For the experimental group the 11 students made 674 errors out of a sample of 2,536 tokens; an error rate of 26.58%. Table 6.19 and Figure 6.4 show the distribution of each tone and tonal errors in the samples of the experimental group for each student. Table 6.20 and Figure 6.5 show the error distribution per tone for the experimental group.

<table>
<thead>
<tr>
<th>Tone</th>
<th>S1</th>
<th>S2</th>
<th>S3</th>
<th>S4</th>
<th>S5</th>
<th>S6</th>
<th>S7</th>
<th>S8</th>
<th>S9</th>
<th>S10</th>
<th>S11</th>
<th>Total</th>
<th>% tones</th>
<th>% Errors</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>481</td>
<td>182</td>
<td>92</td>
<td>139</td>
<td>210</td>
<td>251</td>
<td>183</td>
<td>189</td>
<td>340</td>
<td>307</td>
<td>162</td>
<td>2536</td>
<td>100</td>
<td>26.58</td>
</tr>
<tr>
<td>2</td>
<td>120</td>
<td>61</td>
<td>25</td>
<td>25</td>
<td>38</td>
<td>49</td>
<td>48</td>
<td>37</td>
<td>65</td>
<td>59</td>
<td>34</td>
<td>561</td>
<td>22%</td>
<td>19.61</td>
</tr>
<tr>
<td>3</td>
<td>81</td>
<td>38</td>
<td>16</td>
<td>20</td>
<td>39</td>
<td>46</td>
<td>34</td>
<td>40</td>
<td>63</td>
<td>56</td>
<td>31</td>
<td>464</td>
<td>18%</td>
<td>22.63</td>
</tr>
<tr>
<td>4</td>
<td>151</td>
<td>38</td>
<td>29</td>
<td>54</td>
<td>54</td>
<td>71</td>
<td>20</td>
<td>55</td>
<td>103</td>
<td>93</td>
<td>50</td>
<td>718</td>
<td>28%</td>
<td>30.78</td>
</tr>
</tbody>
</table>

Table 6.19 Distribution of different tones and tonal errors in experimental group’s oral production.
Figure 6.4 Distribution of different tones and tonal errors of experimental group

Table 6.20 Error distribution per tone for the experimental group

<table>
<thead>
<tr>
<th>Subjects</th>
<th>Total tones</th>
<th>Total no. of errors</th>
<th>Total Errors (%)</th>
<th>Tone 1 Errors (%)</th>
<th>Tone 2 errors (%)</th>
<th>Tone 3 errors (%)</th>
<th>Tone 4 errors (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>S1</td>
<td>481</td>
<td>64</td>
<td>13.31</td>
<td>23.44</td>
<td>15.63</td>
<td>18.75</td>
<td>42.19</td>
</tr>
<tr>
<td>S2</td>
<td>182</td>
<td>36</td>
<td>19.78</td>
<td>13.89</td>
<td>0.00</td>
<td>61.11</td>
<td>25.00</td>
</tr>
<tr>
<td>S3</td>
<td>92</td>
<td>17</td>
<td>18.48</td>
<td>5.88</td>
<td>0.00</td>
<td>64.71</td>
<td>29.41</td>
</tr>
<tr>
<td>S4</td>
<td>139</td>
<td>29</td>
<td>20.86</td>
<td>20.69</td>
<td>10.34</td>
<td>55.17</td>
<td>13.79</td>
</tr>
<tr>
<td>S5</td>
<td>210</td>
<td>43</td>
<td>20.48</td>
<td>23.26</td>
<td>0.00</td>
<td>23.26</td>
<td>53.49</td>
</tr>
<tr>
<td>S6</td>
<td>251</td>
<td>41</td>
<td>16.33</td>
<td>31.71</td>
<td>12.20</td>
<td>17.07</td>
<td>39.02</td>
</tr>
<tr>
<td>S7</td>
<td>183</td>
<td>42</td>
<td>22.95</td>
<td>21.43</td>
<td>19.05</td>
<td>9.52</td>
<td>50.00</td>
</tr>
<tr>
<td>S8</td>
<td>189</td>
<td>135</td>
<td>71.43</td>
<td>11.85</td>
<td>20.74</td>
<td>31.85</td>
<td>35.56</td>
</tr>
<tr>
<td>S9</td>
<td>340</td>
<td>136</td>
<td>40.00</td>
<td>11.76</td>
<td>14.71</td>
<td>40.44</td>
<td>33.09</td>
</tr>
<tr>
<td>S10</td>
<td>307</td>
<td>72</td>
<td>23.45</td>
<td>11.11</td>
<td>27.78</td>
<td>30.56</td>
<td>30.56</td>
</tr>
<tr>
<td>S11</td>
<td>162</td>
<td>59</td>
<td>36.42</td>
<td>18.64</td>
<td>18.64</td>
<td>32.20</td>
<td>30.51</td>
</tr>
<tr>
<td>total</td>
<td>2536</td>
<td>674</td>
<td>26.58</td>
<td>16.32</td>
<td>15.58</td>
<td>32.79</td>
<td>35.31</td>
</tr>
</tbody>
</table>
As shown in Table 6.20 and Figure 6.5 shows a total of 2,536 tones were produced by the 11 students in the experimental group. 16.32% of these tones were Tone 1 errors, 15.58% are Tone 2 errors, 32.79% were Tone 3 errors and 35.31% are Tone 4 errors.

Table 6.20 shows the distribution of the different tones and errors made by each subject in the experimental group. Only subject 8 (S8) made a large number of errors which constituted 71.43% of the language produced. On the other hand, subject 1 (S1), for example, produced 481 tones but only around 64 (13.31%) were tonal errors.

### 6.4.3 Order of difficulty of tones

It is difficult to establish an order of difficulty with non-identical samples. However, Table 6.17 and Figure 6.4 show that in the data collected the spread of tones is comparable between experimental and control groups. However, the experimental group produced 1.68 times more language than the control group. The distribution of the four tones for both groups was Tone 4>Tone 3>Tone 1>Tone 2. If the order of difficulty is based on the number of tonal errors
produced, the order of difficulty for both control and experimental groups of students was Tone 4, Tone 3, Tone 1 and then Tone 2 as shown in Figures 6.3 and 6.5.

The order of difficulty of tones obtained differs significantly from other researchers. Table 6.21 shows the order of difficulty obtained by different researchers.

Table 6.21 Order of difficulty for Mandarin Tones by different researchers (including this study) in the field of teaching Mandarin as a FL (CFL) and L1 acquisition by Mandarin children (L1 of Mandarin)

<table>
<thead>
<tr>
<th>Researches</th>
<th>Most difficult</th>
<th>Less difficult</th>
<th>Difficult</th>
<th>Least difficult</th>
<th>Method of testing</th>
<th>Student sample characteristics</th>
</tr>
</thead>
<tbody>
<tr>
<td>Shen (1989b)</td>
<td>Tone 4&gt;</td>
<td>Tone 1&gt;</td>
<td>Tone 3&gt;</td>
<td>Tone 2&gt;</td>
<td>Reading control character text with Pinyin; no questions. All statements.</td>
<td>4 month (1 semester in US), n=8</td>
</tr>
<tr>
<td>CFL</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Miracle (1989)</td>
<td>Tone 2&gt;</td>
<td>Tone 4&gt;</td>
<td>Tone 3=</td>
<td>Tone 1&gt;</td>
<td>Reading: 24 words extracted from statements; only the topics were analysed.</td>
<td>1 year in US, n=10</td>
</tr>
<tr>
<td>CFL</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chen (1997)</td>
<td>Tone 1&gt;</td>
<td>Tone 4&gt;</td>
<td>Tone 2&gt;</td>
<td>Tone 3&gt;</td>
<td>Ask students to make up a narrative using stimulus words. (likely to have no questions)</td>
<td>1 year with different periods of previous study in Mandarin speaking contexts, n=6</td>
</tr>
<tr>
<td>CFL</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>This study</td>
<td>Tone 4&gt;</td>
<td>Tone 3&gt;</td>
<td>Tone 1&gt;</td>
<td>Tone 2&gt;</td>
<td>Naturally occurring speech will a large no. of questions and statements.</td>
<td>30 hours, 5 hours per week for 6 weeks, total zero beginners, n=22</td>
</tr>
<tr>
<td>CFL</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Li &amp; Thompson (1977)</td>
<td>Tone 3</td>
<td>Tone 2</td>
<td>Tone 4</td>
<td>Tone 1</td>
<td>Interview of children with no sentences</td>
<td>between 18-34 months</td>
</tr>
<tr>
<td>Clumeck (1977)</td>
<td>No particular order</td>
<td></td>
<td></td>
<td></td>
<td>Imitation of words with adult models</td>
<td>Child M: 1:2 to 1:11</td>
</tr>
</tbody>
</table>

Key: >: means more difficult than

Table 6.21 shows that many different orders of difficulty for tones have been found by different researchers. An explanation for the different orders is that in some studies the samples of
learners in some studies were not zero beginners. For instance, both Chen and Miracle characterised their learners as having studied Mandarin for 1 year. However, Chen’s sample consisted of three students who had lived in Hong Kong for over ten months; one who had lived in Taiwan for three months and one who lived in Taiwan for over a year and is married to a Chinese woman. Different exposure to Mandarin and tonal language experienced by the students in Chen’s study introduced an unquantifiable variable into the study which has the effect of previous exposure on other tonal languages on the study of Mandarin. Therefore, the different difficulty order for tones could be attributable to prior exposure. If we assume that perception and production are mutually reinforcing processes, then just because some of Chen’s subjects while living in Hong Kong or Taiwan did not speak much Cantonese or Mandarin, it did not mean that they did not perceive the linguistic input in the social environment which would later aid their production in Mandarin (another tonal language). In fact, it could be suggested that long term exposure to such linguistic input would help them develop a “feel” for tones or a “sense of tones”. Therefore, in terms of Chen’s study, these subjects might have developed much further than he had anticipated. In other words, in this case “i” is not the same in Krashen’s “i+1” formula as hypothesised by Krashen (1985). However, in this case “i” is not amenable to quantification or measurement.

Another explanation could be the way data for analyses had been elicited. In the research conducted by Shen and Miracle the methods of elicitation were controlled and accompanied by Pinyin romanization. The texts used also consisted of statements only rather than questions. In Miracle’s study, steps were taken to minimise the effect of sentence intonation on the data obtained. In Chen’s study, he elicited the data more realistically by asking students to generate a
narrative based on a series of stimulus words. In a sense, all three studies tried to minimise the effect of sentence intonation on the data collected.

Similarly in the studies carried out by Clumeck and Li and Thompson in children whose L1 is Mandarin. The input elicited was through imitation of words of one or two syllables rather than sentences. Therefore, the results from L1 acquisition cannot be used to compare with results obtained in teaching Mandarin as a L2 research.

This study is the only study that tried to study a group of homogenous students using a naturalist data elicitation method to collect conversations that are as real as possible. Moreover, because data consisted of conversations on a variety of subject matter such as talking about oneself, one’s family details and hobbies, the data in both groups in this study contained many Tone 3s. Pronouns such as “wō” and “nǐ” were frequently used in the sentence initial position. The data also contained many questions involving the use of “sui” in the sentence final position. Therefore, due to the influence of English question intonation pattern in the final position, errors with Tone 4 were frequent also. There is strong evidence that the beginnings and endings of sentences were a problem area for learners and especially when Tone 4 occurs in the final positions of a sentence or a question. This finding agrees with Wang’s finding that large discrepancies exist between the tonal patterns of Tone 2, Tone 3 and Tone 4 before internal phrase boundaries and those at the end of the utterances exist (Wang, 2001).

Similarly, most of the Tone 3 errors involved either “nǐ: you” or “wō: I” at the beginning of the sentences. For instance, in the experimental group, in Student 1’s data of 571 words, “wō” appeared 24 times. Some “wō” appeared at the sentence initial position in sentences such as “wō jiào Bǐ dé: my name is Peter.” Or “Wō jiā lǐ yǒu bā kǒu rén: there are 8 people in my family”. Others appeared in collocations such as “wō bàba: my father” or “wō māma: my
mother”. When “wǒ: I” occupied the initial position of a sentence but not in the common collocations such as “wǒ bàba: my father,” “wǒ: I” was always pronounced wrongly. For instance in the following figure (Figure 6.6): “wǒ zhù zài Braddon: I live in Braddon”, the NS pronounced “wǒ” as more of a low level tone rather than a low falling tone. The NS’s production of Tone 3 in this position fits in with one of the findings from Miracle’s study (1989). The NSs either produced a falling or level contour in the low register for Tone 3 when in combination with other syllables. This includes when in Tone 3 is in the word initial position of a sentence. Student 1, however, produced “wǒ” as a falling contour similar to Tone 4 in this position. The student’s production of “wǒ” (indicated by the purple line in the highlighted portion) also started a lot higher in register than the NS’s.

Key: Green line represents the teacher’s model. Blue line represents the student’s production. The purple line in the highlighted portion is the student’s rendition of “wǒ” and the brown line is the teacher’s.

Figure 6.6 Mispronunciation of “wǒ zhùzài Braddon.”

Yet, student 1 was able to produce the correct tone contours for “wǒ” in certain common collocations (Figure 6.7) but wrongly in others. In Figure 6.7, student 1 correctly produced “wǒ” as a low register level tone, but he produced “jiējie” incorrectly. This demonstrated the fact that Tone 3 was only partially under control in Student 1’s speech. Again this might
constitutes evidence supporting White’s prediction that English speakers of Mandarin would find it difficult to start a sentence on a high register. They tended to start sentences with a low register.

In short, from a pedagogic point of view, research into order of difficulty of tones does not boost confidence in curriculum designer in designing Mandarin teaching materials. As the pedagogic aim of teaching English L2 learners Mandarin is for them to function intelligibly in Mandarin, it would be wise to abandon the search for the ultimate order of acquisition of Mandarin tones and instead focus on how to solve the problem of L1 interference from English.

6.5 Quality of data produced by both experimental and control groups

In the previous section, differences between the tonal error distribution of the control and experimental groups were reported. However, among the number of errors made by students in the experimental group, the quality of the utterances was different from similar data collected from the control group. In the next section, the data collected from the control and experimental groups will be analysed in terms of its quality.
6.5.1 Quality of oral performances from the control group

Table 6.22 shows the quality of control group’s oral performances in terms of the mean length of utterances, and the number of long sentences produced. Sentences longer than 10 are defined as long sentences. It can be seen the 11 students in the control group produced sentences with a mean length of utterances (MLU) ranging from 2.6 to 8 with an average MLU of 5.7 syllable per utterance.

<table>
<thead>
<tr>
<th>S1</th>
<th>S2</th>
<th>S3</th>
<th>S4</th>
<th>S5</th>
<th>S6</th>
<th>S7</th>
<th>S8</th>
<th>S9</th>
<th>S10</th>
<th>S11</th>
<th>Average</th>
</tr>
</thead>
<tbody>
<tr>
<td>MLU</td>
<td>4.88</td>
<td>2.6</td>
<td>7.1</td>
<td>5.25</td>
<td>5.2</td>
<td>7.32</td>
<td>8</td>
<td>7.92</td>
<td>5.12</td>
<td>4.86</td>
<td>5.7</td>
</tr>
<tr>
<td>No. utterances</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>4</td>
<td>5</td>
<td>2</td>
<td>2</td>
<td>0</td>
<td>1.9</td>
</tr>
</tbody>
</table>

The interaction of the students’ conversations in the control group was mainly in the form of question and answer and no student attempted to incorporate new materials outside the defined scope of the materials taught. Even the longer sentences were from the materials covered in class. For instance, although student 2 (Appendix 7) produced 107 tokens of speech, most of this consisted of answers of no more than 3 syllables in length. In other words, though he fulfilled the requirements of the test, his responses were much shorter than those of others.

6.5.2 Quality of the oral performances of the experimental group

The following Table (Table 6.23) shows the quality of experimental group’s oral performances in terms of the mean length of utterances, the number of long sentences produced, and students’ attempt to expand beyond the language covered in the subject. 11 students in experimental group produced sentences with a mean length of utterances (MLU) ranging from 11.85 (Student 1) to 5.8 (Student 4) with an average MLU of 7.77 for the entire group. Sentences longer than 10 syllables were defined as longer sentences. Every student in the experimental group
produced at least one longer sentence. The whole group produced on average 6.36 long utterances per student compared to the control group. For instance, Student 1 produced 5 longer sentences with the longest one being 61 phonemes long, Student 3 only produced 1 long sentence which was 19 phonemes long. The experimental group produced 3.35 times longer sentences on average than the control group. The interaction pattern demonstrated by the experimental group of students was not simply of a question and answer pattern. Most students set up the conversation context properly using the appropriate language covered. For instance, one conversation was carried out in the context of a visit to a student’s home. Another conversation was carried out in the context of an accidental meeting with a classmate at an airport. The conversations were funny and interesting to listen to for the markers.

<table>
<thead>
<tr>
<th>S 1</th>
<th>S 2</th>
<th>S 3</th>
<th>S 4</th>
<th>S 5</th>
<th>S 6</th>
<th>S 7</th>
<th>S 8</th>
<th>S 9</th>
<th>S 10</th>
<th>S 11</th>
<th>Average</th>
</tr>
</thead>
<tbody>
<tr>
<td>MLU</td>
<td>11.85</td>
<td>5.9</td>
<td>7.0</td>
<td>5.8</td>
<td>7.73</td>
<td>8.94</td>
<td>5.92</td>
<td>3.83</td>
<td>7.91</td>
<td>8.38</td>
<td>12.2</td>
</tr>
<tr>
<td>No. of long utterances</td>
<td>5</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>7</td>
<td>11</td>
<td>4</td>
<td>1</td>
<td>14</td>
<td>14</td>
<td>9</td>
</tr>
</tbody>
</table>

Key: S stands for Student, i.e. S1= Student 1.
Some of the long sentences produced by students contained a number of items in a series and telephone numbers. While the experimental group frequently attempted to expand their spoken materials to include materials outside the normal content scope defined by the classroom materials, the control group did not. The sentence “wǒ yǒu yīge jiējie, jiào Helen, tā sānshí sì suǐ le, tā jiēhunle, yǒu yīge háizi, tā huáiyunle.” was a typical example. This sentence is illustrated above in Figure 6.8 (NS’s rendition: top portion). In the student’s rendition, the ending of each phrase, regardless of the tone, always went up until the last item “Tā huáyùn le” (Figure 6.8 bottom portion). It is visually obvious that at the end of the student’s phrases, the end syllable all had an upward trend, whereas at the end of the same phrases in the NS’s production, the trend for the last syllable was a downward one. Such similar errors were present in the other three students’ data too. This finding lends support to White’s prediction that when speaking on a series of things in Mandarin, English learners of Mandarin tended to pronounce the sequences
of items as a series of rising tones in the last syllable until the last item, which was then pronounced with a falling contour thus replicating the English language intonation employed for the same activity (White, 1981).

6.5.3 Summary on the quality and quantity of the oral performances from both experimental and control groups

The comparison of the quality and quantity of the oral performances from both experimental and control groups discussed in the previous sections demonstrated a number of interesting features:

1. The experimental group produced 1.68 times more language than the control and yet, the experimental group made only slightly more tonal errors than the control group. The experimental group’s error rate was 26.58% of errors and the control group’s error rate was 22%.

2. The order of difficulty for tones was the same for both groups. Tone 4 was the most difficult, followed by Tone 3, Tone 1 and Tone 2.

3. The quality of the oral performances is different with both groups. The experimental group produced longer sentences and more interesting data than the control group. This seems to constitute limited evidence that L2 students in the experimental group seemed to feel more comfortable with the language learned. Consequently, they were willing to appropriate the language as their own more successfully.
6.5.4 Description of Tone errors

6.5.4.1 Tone one – High Level

Because of the uncontrolled nature of the output produced by students, it was difficult to establish NS models as in Miracle’s study (Dreher & Lee, 1968; Guo, 1991). Since the realization of the citation forms of the four tones produced by Miracle’s NS models were in agreement with Guo’s and Dreher and Lee’s findings (Dreher & Lee, 1968; Wang, 2001), Miracle’s NS models of the four tones have been adopted as the standard for the analysis of errors in this study.

According to Miracle’s NS models, the NS model produced a level tone in the upper register of their tonal range (44) in the single tone/syllable words. Before the neutral tone and following Tone Two, Three and Four, this level tone was realized at a lower pitch level (33). Preceding the second tone, this level tone had a tendency towards a falling contour. Acoustic analysis conducted by Wang (2001) showed that at most syllable positions, Tone 1 has an overall decreasing F0 height throughout the utterances. The utterances used were phone numbers containing Tone 1 at nine different positions. The contour errors made by the experimental group consisted primarily of replacing the level contour with a falling contour. 5 students’ production of Tone 1 in their speech samples has been summarized in Table 6.24 for the experimental group. The average range for Tone 1 was 158.07-142.30 Hz with a downtrend of 15.77 Hz. Differences between the beginning of Tone 1 and ending of Tone 1 were between 8.1 Hz to 33 Hz. This tone was almost never produced as a level tone. Though F0 downtrend (please see the Glossary) for Tone 1 was also noted by other researchers (Wang, 2001:74), the differences between the beginnings and ends of Tone 1 were not so great for NSs. The difference was only 4.58 Hz.
Table 6.24 Production of Tone 1 from 5 students in the experimental group

<table>
<thead>
<tr>
<th>Subject</th>
<th>Maximum $F_0$ (Hz)</th>
<th>Minimum $F_0$ (Hz)</th>
<th>Pitch range for Tone 1 (Hz)</th>
<th>Average mean $F_0$ (Hz)</th>
<th>Maximum $F_0$ – average mean $F_0$ (Hz)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>154.80</td>
<td>146.70</td>
<td>-8.1</td>
<td>135.47</td>
<td>19.33</td>
</tr>
<tr>
<td>2</td>
<td>142.17</td>
<td>131.59</td>
<td>-10.58</td>
<td>133.36</td>
<td>8.81</td>
</tr>
<tr>
<td>3</td>
<td>165.00</td>
<td>132.00</td>
<td>-33.00</td>
<td>143.61</td>
<td>21.39</td>
</tr>
<tr>
<td>4</td>
<td>147.05</td>
<td>129.83</td>
<td>-17.23</td>
<td>114.90</td>
<td>32.1</td>
</tr>
<tr>
<td>5</td>
<td>181.34</td>
<td>171.39</td>
<td>-9.95</td>
<td>140.85</td>
<td>40.49</td>
</tr>
<tr>
<td>Average</td>
<td>158.07</td>
<td>142.30</td>
<td>-15.77</td>
<td>133.64</td>
<td>24.43</td>
</tr>
</tbody>
</table>

Key: Maximum $F_0$: Highest point of a subject’s voice. Minimum $F_0$: Lowest point of a subject’s voice. Pitch range: Maximum $F_0$-Minimum $F_0$. Maximum $F_0$-Average mean $F_0$: starting point of Tone 1. -: indicates a downtrend.

Table 6.25 Production of Tone 1 from 2 NSs of Mandarin from the NS group’s data used in the study

<table>
<thead>
<tr>
<th>NS</th>
<th>Maximum $F_0$ (Hz)</th>
<th>Minimum $F_0$ (Hz)</th>
<th>Pitch range for Tone 1 (Hz)</th>
<th>Average mean $F_0$ (Hz)</th>
<th>Maximum $F_0$ – average mean $F_0$ (Hz)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>150.85</td>
<td>147.18</td>
<td>-3.67</td>
<td>128.88</td>
<td>21.97</td>
</tr>
<tr>
<td>2</td>
<td>162.70</td>
<td>157.21</td>
<td>-5.49</td>
<td>145.14</td>
<td>17.56</td>
</tr>
<tr>
<td>Average</td>
<td>156.78</td>
<td>152.20</td>
<td>-4.58</td>
<td>137.01</td>
<td>19.77</td>
</tr>
</tbody>
</table>

Key: Maximum $F_0$: Highest point of a subject’s voice. Minimum $F_0$: Lowest point of a subject’s voice. Pitch range: Maximum $F_0$-Minimum $F_0$. Maximum $F_0$-Average mean $F_0$: starting point of Tone 1. NS: native speaker; -: indicates a downtrend.

A comparison of Table 6.24 and 6.25 shows that L2 learners of Mandarin in the experimental group were all capable of starting Tone 1 at a high enough level except for subject 2 (only 8.81 Hz above his voice range). In fact, the L2 learners started, on average, at a higher frequency (average starting point for Tone 1, 24.43 Hz, see Table 6.24) than NSs (average starting point for Tone 1 was 19.77 Hz, see Table 6.25). Though both the experimental group and the NS group exhibited a downtrend for Tone 1, the downtrend for the experimental group was 15.77 Hz on averaged compared with the 4.58 Hz average for NSs. In other words, it can be concluded that the downtrend in the NSs’ production of Tone 1 did not make them sound non-native because the downturn was small enough not to be deemed wrong perceptually.
However, with the experimental group, the downtrend of 15.77 Hz for Tone 1 was big enough to be deemed wrong perceptually by NSs.

An example of a Tone 1 error is illustrated by the sentence in the Figure 6.9, in which “Tā” was produced as a falling contour.

The register errors consisted of realizing the high level tone too low in the tone space. In Figure 6.29 though the student mispronounced “Tā” with a falling contour as 32, he correctly produced “yī” with a 55 contour. This proves that he was able to start a Tone 1 at the appropriate register. His consistent error with “tā”, at the beginning of an utterance, might lend limited support to White’s prediction that some English speakers would find starting a sentence at a high pitch difficult because in English, the first word or part of the sentence is usually used for showing the listener what the focus of the sentence is. Therefore, the first word or first phrase is likely to
be stressed if it is acting as a focus. As stress in English is usually accomplished by a dip in
frequency, a downward contour for the first word is likely to be produced.

<table>
<thead>
<tr>
<th>Subject</th>
<th>Maximum $F_0$ (Hz)</th>
<th>Minimum $F_0$ (Hz)</th>
<th>Pitch range for Tone 1 (Hz)</th>
<th>Average mean $F_0$ (Hz)</th>
<th>Maximum $F_0$ – average mean $F_0$ (Hz)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Subject 1</td>
<td>155.25</td>
<td>144.38</td>
<td>-10.87</td>
<td>117.25</td>
<td>38</td>
</tr>
<tr>
<td>Subject 2</td>
<td>124.31</td>
<td>113.39</td>
<td>-10.93</td>
<td>114.27</td>
<td>10.04</td>
</tr>
<tr>
<td>Subject 3</td>
<td>110.90</td>
<td>116.94</td>
<td>-6.04</td>
<td>104.88</td>
<td>6.02</td>
</tr>
<tr>
<td>Subject 4</td>
<td>167.33</td>
<td>139.80</td>
<td>-27.53</td>
<td>138.25</td>
<td>29.08</td>
</tr>
<tr>
<td>Average</td>
<td>139.45</td>
<td>128.63</td>
<td>-12.33</td>
<td>118.66</td>
<td>20.79</td>
</tr>
</tbody>
</table>

Key: Maximum $F_0$: Highest point of a subject’s voice.
Minimum $F_0$: Lowest point of a subject’s voice.
Pitch range: Maximum $F_0$ - Minimum $F_0$.
Maximum $F_0$: Average mean $F_0$: starting point of Tone 1.
-: indicates a downtrend.

Four students’ samples from the control group were analysed with respect to the frequency
ranges for the different tones. The control group students made similar errors. Though most
students started Tone 1 above the average mean pitch in the control group, they did not start
sufficiently high enough above the mean pitch. For instance, subject 3 only started Tone 1 at
6.02 Hz above his average mean for his $F_0$ (see Table 6.26) while the two NSs consistently start
Tone 1 at 19.77 Hz above their average mean for their $F_0$ (see Table 6.25).

6.5.4.2 Tone two – Rising

According to Miracle’s NS models, the NSs produced this tone beginning in the low register of
tone space and rising into the upper register of the tone space (24). Dreher and Lee’s (1968)
findings support this description of the second tone. This contour was found in single
tone/syllable words, in word initial position before the neutral, fourth and other second tones
and following third tones. In word final position (except following third tone) and before the
first and third tones, this tone was realized as level contour in the lower register of the tone
space (22 or 23). Preceding the first tone, this level contour was found in the mid register. Wang (2001) investigated the behaviour of tones in different types of utterances and the effect of phrase boundary on tone realization. It was found that “Tone 2 has an overall rising shape; however, there is a small falling portion preceding a larger rise at non-phrase-initial positions” (Wang, 2001). In Table 6.27, NSs produced Tone 2 consistently with a rising contour and the starting points of Tone 2 for each speaker were below their average mean F₀ for their voice (starting at 16 Hz below their average mean F₀).

<table>
<thead>
<tr>
<th>Tone 2</th>
<th>Maximum F₀ (Hz)</th>
<th>Minimum F₀ (Hz)</th>
<th>Pitch range for Tone 1 (Hz)</th>
<th>Average mean F₀ (Hz)</th>
<th>Maximum F₀ – average mean F₀ (Hz)</th>
</tr>
</thead>
<tbody>
<tr>
<td>NS 1</td>
<td>116.79</td>
<td>133.62</td>
<td>+16.82</td>
<td>128.88</td>
<td>+12.09</td>
</tr>
<tr>
<td>NS 2</td>
<td>125.24</td>
<td>150.83</td>
<td>+25.59</td>
<td>145.14</td>
<td>+19.9</td>
</tr>
<tr>
<td>Average</td>
<td>121.02</td>
<td>142.23</td>
<td>+21.21</td>
<td>137.01</td>
<td>+16</td>
</tr>
</tbody>
</table>

Key:

“-”: downtrend (going down from the beginning of the syllable to the end);

“+”: uptrend (going up from the beginning of the syllable to the end)

NS: native speaker

Maximum F₀: Highest point of a subject’s voice.

Minimum F₀: Lowest point of a subject’s voice.

Pitch range: Maximum F₀ - Minimum F₀.

Maximum F₀ - Average mean F₀: starting point of Tone 2.

The data on Tone 2 from both the control and experimental groups of students differ from NSs’ production in a fundamental way. The register mistakes consisted primarily of beginning tone too high in the tone space (see Student 1 in Table 6.27). Student 1 started the beginning of Tone 1 at 172.41 Hz which was 36.94 Hz above his average mean F₀. He also consistently produced Tone 2 with a downtrend rather than uptrend.
Table 6.28 Students’ production of Tone 2 from the experimental group

<table>
<thead>
<tr>
<th>Tone 2</th>
<th>Maximum $F_0$ (Hz)</th>
<th>Minimum $F_0$ (Hz)</th>
<th>Pitch range for Tone 1 (Hz)</th>
<th>Average mean $F_0$ (Hz)</th>
<th>Maximum $F_0$ – average mean $F_0$ (Hz)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Subject 1</td>
<td>172.41</td>
<td>146.30</td>
<td>-26.11</td>
<td>135.47</td>
<td>-36.94</td>
</tr>
<tr>
<td>Subject 2</td>
<td>122.81</td>
<td>144.49</td>
<td>+21.68</td>
<td>133.36</td>
<td>+10.55</td>
</tr>
<tr>
<td>Subject 3</td>
<td>107.50</td>
<td>190.75</td>
<td>+83.25</td>
<td>143.61</td>
<td>+36.11</td>
</tr>
<tr>
<td>Subject 4</td>
<td>114.81</td>
<td>120.74</td>
<td>+5.93</td>
<td>114.90</td>
<td>+0.09</td>
</tr>
<tr>
<td>Subject 5</td>
<td>123.99</td>
<td>144.59</td>
<td>+20.60</td>
<td>140.85</td>
<td>+16.86</td>
</tr>
<tr>
<td>average</td>
<td>128.30</td>
<td>149.37</td>
<td>+21.07</td>
<td>133.64</td>
<td>-5.34</td>
</tr>
</tbody>
</table>

Key:

“-”: downtrend (going down from the beginning of the syllable to the end);
“+”: uptrend (going up from the beginning of the syllable to the end)

Maximum $F_0$: Highest point of a subject’s voice.
Minimum $F_0$: Lowest point of a subject’s voice.
Pitch range: Maximum $F_0$-Minimum $F_0$.
Maximum $F_0$-Average mean $F_0$: starting point of Tone 2.

The contour mistakes came from substituting either a falling or level contour for the desired rising contour. While most students managed to produce Tone 2 by producing them as the desired rising contour, some students produced them consistently with a falling contour or a flat contour (Student 1 and 4). A similar trend can be observed from students’ production of Tone 2 in the control group as shown in Table 6.29 below.

Table 6.29 Students’ production of Tone 2 from the control group

<table>
<thead>
<tr>
<th>Tone 2</th>
<th>Maximum $F_0$ (Hz)</th>
<th>Minimum $F_0$ (Hz)</th>
<th>Pitch range for Tone 1 (Hz)</th>
<th>Average mean $F_0$ (Hz)</th>
<th>Maximum $F_0$ – average mean $F_0$ (Hz)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Subject 1</td>
<td>111.50</td>
<td>134.75</td>
<td>+23.25</td>
<td>117.25</td>
<td>+5.75</td>
</tr>
<tr>
<td>Subject 2</td>
<td>132.38</td>
<td>155.70</td>
<td>+23.33</td>
<td>143.27</td>
<td>-18.11</td>
</tr>
<tr>
<td>Subject 3</td>
<td>111.29</td>
<td>107.09</td>
<td>-4.20</td>
<td>104.88</td>
<td>-6.41</td>
</tr>
<tr>
<td>Subject 4</td>
<td>116.19</td>
<td>142.99</td>
<td>+26.80</td>
<td>138.25</td>
<td>+22.06</td>
</tr>
<tr>
<td>average</td>
<td>117.84</td>
<td>135.13</td>
<td>+19.40</td>
<td>133.66</td>
<td>+0.82</td>
</tr>
</tbody>
</table>

Key: “-”: downtrend (going down from the beginning of the syllable to the end);
“+”: uptrend (going up from the beginning of the syllable to the end)

Maximum $F_0$: Highest point of a subject’s voice.
Minimum $F_0$: Lowest point of a subject’s voice.
Pitch range: Maximum $F_0$-Minimum $F_0$.
Maximum $F_0$-Average mean $F_0$: starting point of Tone 2.
In Figure 6.10, the student produced “cóng” as “còng”. Fewer errors were produced in Tone 2 because in this tone is also realized as a level contour in some environments in the NS models. For instance, all the students in the experimental group were able to produce “shí” correctly when it is in the middle of two numbers such as in “wūshíyī” (Figure 6.11).

Key: Green line represents the teacher’s model. Blue line represents the student’s production.

Figure 6.10. Mispronunciation of “Cóng” in “Colin cónɡ England lái de”

6.5.4.3 Tone three - Low Dipping

According to Miracle’s (1989) NS samples, the NSs produced a low register dipping tone (212) in single tone/syllable words and either a falling or level contour in the low register when in combination with other syllables. When in the word initial position, the low level contour was more prevalent (22); while in the word final position, the low falling contour (21) was more prevalent. A consistently falling contour was found preceding the neutral tone (21) and Tone
One (21). In all other combination environments there was variation among the NSs between a falling and level contour.

However, complicated tone sandhi rules for Tone 3 meant that mistakes were frequently made with this tone. The contour errors made by the students consisted primarily of realizing the Tone 3 as a rising contour or replacing a level tone with a falling or dipping contour (Figure 6.11 with wō). The confusion between the falling and level contours is clearly understandable given this tone’s various realizations by the NSs.

Key: Green line represents the teacher’s model. Blue line represents the student’s production.

Figure 6.11 Mispronunciation of “wō” in “wō jiào X”.

Furthermore, as expected, students consistently produced “wō” before another Tone 3 wrongly (Figure 6.12).
Key: Green line represents the teacher’s model. Blue line represents the student’s production.

Figure 6.12 Mispronunciation of “wōhên” in “rènshī nǐ, wō hēn gāoxǐng”.

The tone sandhi rule for Tone 3 states that when a Tone 3 is before another Tone 3, the first Tone 3 becomes Tone 2, and the second one preserves the 21 contour, students’ production consistently produced the first Tone 3 as the normal falling contour (in this case almost flat).

The register errors consisted of realizing Tone 3 too high in the tone space (Figure 6.11).

6.5.4.4 Tone four – High Falling

Miracle’s (1989) NS models showed the NSs were consistent in producing a falling contour tone beginning in the upper register of the tone space and falling into the lower tone register (51) for Tone Four in the single syllable/tone words. In the tone combination words, this steep drop was also found before Tone 3, the neutral tone and following Tone 2. Preceding Tones 1, 2, and 4 the decline is not so great (53) and following Tones 1, 3, and 4 the falling contour starts lower in the tone space and does not drop as far (42).
Students from both control and experimental groups realized most of the Tone 4 as the expected downward contour. However, two common problems remained compared with NS data in Table 6.32. One is that students did not start Tone 4 high enough in the tone space. While most of the students in the experimental group started Tone 4 above their average mean F₀, student 3 sometimes started above, sometimes below the mean F₀ of his voice.

Second, though a downward contour was produced consistently for all the subjects, students did not go low enough with Tone 4. In Table 6.32, the difference between the beginning of all Tone 4 and the ends of all the Tone 4 produced by the NSs was 37.69 Hz downward. Table 6.30 and 6.31 show that for the experimental and control groups of students, the differences were only 20.26 Hz and 26.23 Hz respectively.

<table>
<thead>
<tr>
<th>Tone 4</th>
<th>Subject 1</th>
<th>Subject 2</th>
<th>Subject 3</th>
<th>Subject 4</th>
<th>Subject 5</th>
<th>Average</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Maximum F₀ (Hz)</td>
<td>Minimum F₀ (Hz)</td>
<td>Pitch range for Tone 4 (Hz)</td>
<td>Average mean F₀ (Hz)</td>
<td>Maximum F₀ – average mean F₀ (Hz)</td>
<td></td>
</tr>
<tr>
<td>Tone 4</td>
<td>152.35</td>
<td>113.71</td>
<td>-38.63</td>
<td>135.47</td>
<td>-16.88</td>
<td></td>
</tr>
<tr>
<td></td>
<td>142.17</td>
<td>131.59</td>
<td>-10.58</td>
<td>133.36</td>
<td>-8.81</td>
<td></td>
</tr>
<tr>
<td></td>
<td>131.33</td>
<td>108.17</td>
<td>-23.17</td>
<td>143.61</td>
<td>+12.28</td>
<td></td>
</tr>
<tr>
<td></td>
<td>141.60</td>
<td>122.28</td>
<td>-19.32</td>
<td>114.90</td>
<td>-26.7</td>
<td></td>
</tr>
<tr>
<td></td>
<td>153.43</td>
<td>123.57</td>
<td>-29.86</td>
<td>140.85</td>
<td>-12.58</td>
<td></td>
</tr>
<tr>
<td></td>
<td>120.81</td>
<td>99.89</td>
<td>20.92</td>
<td>133.64</td>
<td>+12.83</td>
<td></td>
</tr>
</tbody>
</table>

Key: “-”: downtrend (going down from the beginning of the syllable to the end);
“+”: uptrend (going up from the beginning of the syllable to the end)

Maximum F₀: Highest point of a subject’s voice.
Minimum F₀: Lowest point of a subject’s voice.
Pitch range: Maximum F₀-Minimum F₀.
Maximum F₀-Average mean F₀: starting point of Tone 4.
### Table 6.31 Students’ production of Tone 4 from the control group

<table>
<thead>
<tr>
<th>Subject</th>
<th>Maximum $F_0$ (Hz)</th>
<th>Minimum $F_0$ (Hz)</th>
<th>Pitch range for Tone 4 (Hz)</th>
<th>Average mean $F_0$ (Hz)</th>
<th>Maximum $F_0$ – average mean $F_0$ (Hz)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Subject 1</td>
<td>129.56</td>
<td>109.00</td>
<td>-20.56</td>
<td>117.25</td>
<td>-12.31</td>
</tr>
<tr>
<td>Subject 2</td>
<td>128.14</td>
<td>109.60</td>
<td>-18.54</td>
<td>114.27</td>
<td>-13.87</td>
</tr>
<tr>
<td>Subject 3</td>
<td>113.79</td>
<td>107.44</td>
<td>-6.35</td>
<td>104.88</td>
<td>-8.91</td>
</tr>
<tr>
<td>Subject 4</td>
<td>167.97</td>
<td>108.50</td>
<td>-59.47</td>
<td>138.25</td>
<td>-29.72</td>
</tr>
<tr>
<td>Average</td>
<td>134.87</td>
<td>108.64</td>
<td>-26.23</td>
<td>118.66</td>
<td>-16.21</td>
</tr>
</tbody>
</table>

Key: “-“: downtrend (going down from the beginning of the syllable to the end);
“+” : uptrend (going up from the beginning of the syllable to the end)

Maximum $F_0$: Highest point of a subject’s voice.
Minimum $F_0$: Lowest point of a subject’s voice.
Pitch range: Maximum $F_0$-Minimum $F_0$.
Maximum $F_0$-Average mean $F_0$: starting point of Tone 4.

### Table 6.32 NSs’ production of Tone 4

<table>
<thead>
<tr>
<th>NS</th>
<th>Maximum $F_0$ (Hz)</th>
<th>Minimum $F_0$ (Hz)</th>
<th>Pitch range for Tone 4 (Hz)</th>
<th>Average mean $F_0$ (Hz)</th>
<th>Maximum $F_0$ – average mean $F_0$ (Hz)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ns1</td>
<td>162.01</td>
<td>119.98</td>
<td>-42.03</td>
<td>128.88</td>
<td>-33.13</td>
</tr>
<tr>
<td>Ns2</td>
<td>160.49</td>
<td>127.31</td>
<td>-33.35</td>
<td>145.14</td>
<td>-15.35</td>
</tr>
<tr>
<td>Average</td>
<td>161.25</td>
<td>123.65</td>
<td>-37.69</td>
<td>137.01</td>
<td>-24.24</td>
</tr>
</tbody>
</table>

Key: “-“: downtrend (going down from the beginning of the syllable to the end);
“+” : uptrend (going up from the beginning of the syllable to the end)

Maximum $F_0$: Highest point of a subject’s voice.
Minimum $F_0$: Lowest point of a subject’s voice.
Pitch range: Maximum $F_0$-Minimum $F_0$.
Maximum $F_0$-Average mean $F_0$: starting point of Tone 4.

With Tone 4, most of the errors were made in the sentence final position in sentences such as “X duoldàle?” (How old is X?). Such errors were usually a result of transferring English intonation pattern of a question, which goes up at the end, for a question in Mandarin. In the sentence “tāduōdàle?: How old is he?” (Figure 6.13), the pitch contours indicated in the following graph shows the student’s production (Green line) goes up rather than down like in the teacher’s model (Blue line).
Tonal values of Tone 5 (neutral tone) depend on the tonal value of the preceding tone and the sentence intonation. Pedagogically, syllables with neutral tones only occur in a small subset of two or three syllable compounds but in sentences, neutral tones also are affected by sentence intonation. So teaching and offering feedback to the neutral tone are difficult. Therefore, in this thesis, it was decided not to compute them. In this study, it has been observed that neutral tones in words such as “bāba: father” and “māma: mother” were produced correctly but errors occur at the end of sentences with sentence types such as questions.
6.6 Types of Errors

6.6.1 Interference Errors

One of the influential papers on predicting the interference errors from English intonation were Caryn White’s paper titled “Tonal pronunciation errors and interference from English Intonation”. In this paper, White (1981) used contrastive analysis to predict errors that are likely to be made by learners of Mandarin whose native language is American English. The primary assumption adopted by White is that perception necessarily precedes production and therefore assumes that once “correct” perception of the tones has been achieved, correct production will naturally follow. Using contrastive analysis and with this assumption in mind, White exhaustively described in detail how English speakers might transfer English intonation patterns onto Mandarin sentences. She summarized that an English speaker of Mandarin might have trouble in the following areas (White, 1981):

1. It might be difficult for an English speaker to begin an utterance on a rising tone (that is Tone 2 or Tone 1 or Tone 4) because so few English utterances begin this way.

The results from both the control and experimental groups show that many English L2 learners of Mandarin do have a problem starting a sentence at a high level at the sentence initial level. For instance, most students do not have a problem with Tone 1 if it is within a sentence. It is at the beginning of a sentence that they have trouble with. The cause of this type of errors is most likely to come from the unconscious transfer of English intonation and discourse patterns into Mandarin production.

2. In uttering sequences of items, items in the sequence are usually pronounced as a series of rising tones until the last item, which is pronounced with a falling contour.
Unfortunately, the language samples from the control group were limited in their length of utterances and interaction patterns. No student managed to produce any sentences longer than one single statement or question. This means that no proof of this type of language transfer can be obtained from this data. However, the subjects from the experimental group did produce much longer utterances. In sentences involving a series of items, this group of learners did pronounce the language items with a rising tone except the last item which was usually pronounced with a downward contour. This clearly was due to the interference of English intonation into Mandarin.

3. Errors in pauses might be related to L1 transfer. In English, word boundaries are marked not by pauses but by phonological features and/or stress. But in Mandarin it is the potential for a pause to distinguish two monosyllabic words from one disyllabic word. Therefore in recalling a disyllabic word, if a pause is introduced in between the two monosyllabic components, the production of the disyllabic word would be unnatural and sometimes meaning can be altered.

The data from the control and experimental groups did not contain any error data related to pausing or errors in segmenting disyllabic words or trisyllabic words wrongly. Every student, regardless of their language ability, produced sentences with the appropriate word boundaries. While many students had problems with “wǒ” by itself in front of a verb such as “zhù”, most of them produced collocations such as “wǒ māma” or “wǒ bàba” correctly. This indicates that by teaching the utterances in their context without cutting them up with rhythm and stress patterns preserved, students intuitively pick up the grouping of words with the appropriate pauses and stress.
4. White (1981) predicted that emphatic stress, associated with a high-falling tone in English, can cause problems for beginning Mandarin students perhaps because of its similarity to the Mandarin falling tone. This is particularly noticeable in the case of descriptive adjectives that precede the noun being described. White predicted that if it happens that the Mandarin word has a falling tone, such as “dà”, “big, large”, then it is seldom mispronounced by English speakers. However, in the case of the host of adjectives that are not pronounced with the falling tone in Mandarin, they are often mispronounced as Tone 4 by speakers of English. For example, the English speakers wishing to emphasize the Mandarin word “hěn” “very” in “hěn hǎo” may say: “hènhǎo”. Unfortunately, as the language data produced did not contain a large number of descriptive phrases, no evidence supporting this prediction can be found in the present sample.

5. Beginning English speakers of Mandarin are most likely to be influenced by the intonation patterns of English when speaking Mandarin. This is because intonation functions on a deeper level of consciousness that makes the transfer of intonation from L1 into L2 during language production inevitable.

Many errors in the data indicated a strong tendency for English learners of Mandarin to transfer the intonation patterns of English into their learning of L2. This is most evident at the beginning and the end of the utterances.

Another area of difficulty experienced by English learners of Mandarin is in terms of register (Miracle 1989; Shen 1989). Research findings from Chen (1997) concurred with Miracle and Shen that American learners of Mandarin tend to articulate Mandarin sounds within the 2-4 level of their voice range. The groups of students involved in this study continued to have
register problems. Because of the difference in English speakers’ voice range from that of the NSs of Mandarin, the inability for English speaker learners to start Tone 1 and Tone 4 at a high enough level and the inability for them to go lower than the lowest limit of their voice range for Tone 4 meant that even though the contour of the tones might be kept, the problem with register persisted.

### 6.6.2 Induced errors

The teaching methodology did not involve analyzing particular tones by its tone name; rather it emphasized the importance or rhythm, stress and sentence intonation. Therefore, if there were any induced errors, it would have to do with how the Tone 3 tone sandhi phenomenon was not treated. Researchers and teachers have devoted much time and energy to discovering the best teaching method for teaching Tone 3. Lin (1985) and recently Hsieh (1996) correctly suggested that Tone 3 should be taught as a low-dipping because in running discourse, Tone 3 is usually realised as low-dipping. Pedagogically, the behaviour of Tone 3 due to its complex tone sandhi rules is too complex to be taught explicitly without at the same time providing students with ample opportunities for perceptual training with this tone.

The perceptual training could consist of using the materials covered in this study as an example. Firstly group all the sentences starting with “wǒ” together and then ask the students to observe the behaviour of Tone 3 in these utterances. Out of 104 sentences, 47 sentences contained Tone 3 in various combinations. Out of the 14 sentences in Week 1’s materials, 10 sentences had Tone 3 in them. Therefore, the need to sensitize learners to the behaviour of Tone 3 in various combinations is important.
6.7 Summary

A major aim of this chapter is to summarize quantitative information sought in the following areas:

1. Sequence of mastery of the four tones, and hence the order of difficulty of those tones;
2. Common tone and prosodic error patterns among learners;
3. Likely cause(s) leading to the detected errors;
4. How and where interference from English intonation takes place and how significant is its role;
5. Effectiveness of SEA on the teaching of Mandarin tones and prosody;

Findings of this study suggest that SEA influenced the acquisition of Mandarin prosody and tones of the experimental group in a profound way. Acoustically, after only six weeks (30 Hours) of training, students taught by SEA exhibited behavioural changes in the way they spoke Mandarin. Their voice quality changed profoundly when speaking Mandarin compared to when they spoke English, their mother tongue. In other words, when speaking Mandarin, they spoke with a higher average mean fundamental frequency ($F_0$) with wider pitch range (measured by the average mean $F_0$ SD) and were capable of raising the maximum of their voice range to accommodate for tones. After only 30 hours of training, these students behaved more like NSs of Mandarin in terms of their voice range and possibly in other areas as well. For instance, though not measured in this study, it is entirely probable that when speaking Mandarin, learners might also have changed the way they carry their bodies and other non-verbal behaviour patterns.

These objectively measured changes in experimental group students’ voice quality were then verified by a group of nine NSs of Mandarin through a perceptual experiment. Subjective
perceptual results of this experiment confirm that the students taught by SEA from the experimental group performed better than the control group who were not taught by SEA. Furthermore, the level of agreement reached by the nine NSs was very high as indicated by an inter-rater reliability score of 0.92.

Unfortunately, the acoustic and perceptual experiments were not able to inform us more on the nature of the prosodic errors made by L2 learners of Mandarin. Thus, an auditory test was also performed by three experienced teachers of Mandarin to explore in detail the nature of the errors made by these students.

It was found that though the students in the experimental and control groups behaved very differently in terms of their voice quality, common error and acquisition patterns do exist:

1. Firstly, for both groups, the order of difficulty of tones was the same.
2. Secondly, both groups of students had great trouble with the beginning and end of utterances.
3. Thirdly, both groups of students did not have any trouble with pausing and knowing where the sentence boundaries were.

The common patterns of errors could be traced back to the interference of English discourse and intonation patterns in speaking Mandarin. However, the large number of errors in Tone 3 could be caused by ineffective measures taken to teach Tone 3. The teaching of Tone 3 could be enhanced by grouping a large number of utterances with Tone 3 together so that students can observe the different realizations of Tone 3 at their own pace. Sptool designed for this study could have been used in the teaching of Tone 3.
In the next chapter, the qualitative results of both the control and experimental groups of students are discussed. The qualitative results obtained from questionnaires administered at the end of the course and face to face interviews will provide yet another source of information on the efficacy of SEA for English L2 learners of Mandarin.
Chapter 7 Qualitative results for both control and experimental groups of students

Chapter 6 presented quantitative results for both control and experimental groups of students in terms of physical changes in voice ranges and tonal errors made in spoken performances. This chapter presents the qualitative results of both groups through a description and analysis of the data for each group of students. An analysis of the data for each group is presented and related to the students’ Mandarin language learning practices.

The instruments used to collect the qualitative data were an end of course questionnaire and a face-to-face interview with each participating student. The end of course questionnaire was designed to obtain data on how students used various technology provided in the course. Biographical information about each student was also gathered in the end of course questionnaires.

The information on the student’s cognitive processes was gathered from an interview with each student at the end of the semester. Although a set of guiding of questions were prepared as reference for the interviewer (see Appendix 4), the students were encouraged to talk widely and freely. Each interview lasted for about 45 minutes. The aim of the interview was to elicit retrospective data about the strategies used in learning Mandarin in the first 13 weeks of their Mandarin language learning. Care was taken to frame open-ended questions so that the student responses would not be affected by the interview situation. Following the advice of Field and Morse (1985), attention was given to ensure a relationship of trust. The interviews were transcribed word for word, as suggested by Field and Morse (1985). The results of the questionnaires and the interviews from the control and experimental groups will be discussed separately.
7.1 Control group

7.1.1 Results of end of course questionnaires

A questionnaire about the course materials was distributed to every student at the end of the semester (see Appendix 8). Twenty one students completed the subject and the questionnaires included non-participants of the experiments.

From the responses of questionnaire, 85% of the students regularly used the audio and dataCD for at least 1 to 3 hours each week. Students who did not use the CDs were Thai exchange students who did not have access to either a CD player or computer and consequently could not make use of the CDs. The entire class stated that such technology-based materials were extremely useful to their studies.

7.1.2 Interview data

At the end of second semester in 2001, a face-to-face interview was conducted between the researcher and each student in the control group. Unfortunately, by the time the interviews were conducted, many Australian students who were the original participants in the control group had already left the course. Consequently, interview data from only 4 of originally participating students were collected. Because of the small number of interviews conducted, the analysis would mainly contain extracts from the data to illustrate certain emergent trends. Out of the control group, interview data were available for students S1, S2, S5 and S6.

The categorization of strategies used to analyse this data is based on those of Oxford (1989) (see Appendix 5). However, the categorization of strategies was not taught explicitly to either group of students at the beginning of the first semester. In other words, they were not used as
prescriptive tools for strategy change. Nevertheless, it was useful to use the categorization to interpret the data obtained in this study.

7.1.2.1 The control group’s use of cognitive strategies in language learning

7.1.2.1.1 Direct strategy use of the control group

Table 7.1 Direct (performance) strategy use of four students in the control group.

<table>
<thead>
<tr>
<th>Students</th>
<th>Creating mental images</th>
<th>Applying images and sounds</th>
<th>Reviewing well</th>
<th>Employing action</th>
<th>Practicing</th>
<th>Receiving and sending messages</th>
<th>Analyzing and reasoning</th>
<th>Creating structure for input and output</th>
<th>Guessing intelligently</th>
<th>Overcoming limitations in speaking and writing</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>6</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>14</td>
</tr>
<tr>
<td>2</td>
<td>3</td>
<td>4</td>
<td>1</td>
<td>0</td>
<td>7</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>15</td>
</tr>
<tr>
<td>5</td>
<td>3</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>10</td>
</tr>
<tr>
<td>6</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>4</td>
</tr>
<tr>
<td>total</td>
<td>12</td>
<td>8</td>
<td>3</td>
<td>3</td>
<td>13</td>
<td>1</td>
<td>2</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>43</td>
</tr>
</tbody>
</table>

Table 7.2 Indirect strategy use of four students in the control group.

<table>
<thead>
<tr>
<th>Students</th>
<th>Evaluating your learning</th>
<th>Lowering your anxiety</th>
<th>Arranging and planning your learning</th>
<th>Asking questions</th>
<th>Cooperating with others</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>5</td>
<td>0</td>
<td>5</td>
<td>0</td>
<td>7</td>
</tr>
<tr>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>5</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>6</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>total</td>
<td>1</td>
<td>5</td>
<td>2</td>
<td>8</td>
<td>1</td>
<td>12</td>
</tr>
</tbody>
</table>

Table 7.1 shows that students in the control group used many strategies to improve their memory. However, these memorization strategies primarily concern the memorization of characters. In the students’ interviews, these students stated that they considered having to memorize language items as the biggest obstacle in language learning. As Student 1 commented in the Interview:

I learned Japanese as well. In Japanese is “hiragana”, which is how you basically pronounce things but you’d never seen the Chinese character before, [you] don’t know how it is really pronounced even sometime [they] have the phonetic elements but they are not always the same. Or you might know it, or you don't know which tones. So, you just have to memorize each individual character. That’s why it’s such a huge thing…You have to increase your vocabulary the whole time.
Memorization is a big part of learning languages in general but it is extremely important in the learning of Mandarin characters especially when one has to learn the written script at the same time as the spoken language. As Student 1 pointed out rightly that in memorising an individual character, for a Mandarin speaking child, three aspects of the character have to be remembered for it to be meaningful: that is the way the character is written; how to pronounce it and its meaning. For a non native learner of Mandarin, in the traditional teaching pedagogy, the learning of the tone of the character has also been added to the other three aspects of the character. For a Mandarin speaking child, though the spoken language is usually learned within the first three years of his or her life, the mastery of the written language takes around 6 years to master. Learning the 3000 Chinese characters which makes one capable of reading a Chinese newspaper involves writing around 20 characters per day everyday throughout a child’s primary school life. We also need to remember that a Chinese school child learns these characters not in a vacuum but in an environment rich with characters. Thus the motivation for mastering Chinese characters is enormous.

The situation of learning Mandarin as a FL is very different from that of a Mandarin speaking child. Firstly, the language is not taught everyday. In the present study, the language was only taught for five hours per day. Secondly, students tended not to be surrounded by an environment rich with characters. Thirdly, in this study they were only required to learn around 20 characters per week. In other words, the learning situation is not motivating for learning Chinese characters. One other fundamental difference is that learning to write Chinese characters as a Mandarin speaking child involves training new neural pathways to form in the brain which leads to new motor skills involved in writing characters. In L2 Mandarin students, the training is perhaps not just to form new neural pathways in the brain to write characters but to alter pre-
existing neural pathways which have been used to form English letters. If a phonological sieve exists for phonemes of a language, it would be reasonable to suggest that a “sieve” also exists for learning to write in a foreign script or any other aspects of language learning.

Respondents in the interview in the control group expressed difficulty in memorizing Chinese characters and they seemed to solve the issue by either creating mental images of the characters or by writing the characters over and over again. In the interviews, the following comments were some students’ description of how they made associations between sounds and characters:

Student 1:

[I] memorise characters and do character work. And in my head I try, like I write the characters and I try to associate with the concept that they were in instead of just English translation because that way I don’t just translate into English.

Student 2:

There is definitely the image thing going on. Not necessarily the shape of characters, but senses, features, drawings, that kind of thing. That’s will really help in setting the scene for grabbing the memory at a particular location.

Student 5:

I try to remember the characters. Definitely, sounds and phrases have to be in characters.

For these students, imagery was used to relate to the characters. They did not use imagery to memorize any other aspects of the language. For instance, students in the control group seemed to employ fewer strategies to practice the language. In the interviews, the following was how some students described their practising behaviour:

Student 2:
The regular character tests for this semester were good.

I just wrote out the characters again and again and every week the test was good.

Just immersion, writing the characters again and again and again.

Student 5:

I think the character course is really good because there is motivation to know the characters each week.

As it can be seen from the interviews, the focus again was on the writing and retention of characters. Only one student commented on the importance of practicing speaking and listening. This same student also tried to use analytical skills (Student 1) and another used Compensatory skills (Student 6).

In short, students in the Control group primarily saw the learning of Mandarin as learning the written language. They seemed to focus on the difficulty of the language rather than their achievement.

7.1.2.1.2 Indirect strategy use of the control group

In terms of “Indirect” strategies, two students had regular contacts with NSs (Student 1 and Student 2). For these students, the social contacts they had with NSs were useful. NS contacts were either for immersion purposes or for learning the language in context or to obtain feedback:

Student 1:

I have a Chinese friend. He can speak a bit of Mandarin as well. So, when I was starting to try to learn tones. He always corrects me and shows me how to do it that sort of thing. He always corrects me, a lot of time. And I just got to mimic him, what he does. And after little while, you get the hang of what’s going on and so you can sort of do itself a bit more.
And then I just try to do sentences and stuff. I just show my friends sentences in Chinese to see if they work because I don’t know how to write sentences in English. I mean in Chinese because I try sometime because it’s similar in English and Chinese. That I try to do things we do in English with sentences but sometime they don’t work.

Student 2:

Immersion idea. As I have said before. Going over it again and again. I find that’s really useful. And also hearing a different accent. Also having different tutors during the semester has been helpful as well.

However, only Student 1 used strategies to lower his anxiety:

Student 1:

I think tones link to music. I listen to the music of the tones….. I think I feel like in my head, I just think it is sort of music.

Because my friend whose husband is Chinese, [we] play Japanese games together.

Student 1 was also the only one who watched movies to practice listening in Mandarin. Though the control group students seemed to be pre-occupied with Chinese characters, in terms of memorizing the spoken language, all the students tended to recall phrases and sentences learned in their contexts rather than just individual words or character:

Student 2:

Some of them I am remembering in isolation, some of them I am remembering the part of the phrase.

Student 5:

I like phrase Chinese very much. I like “nín guixing?”, that kind of phrase. Not just one character, but phrases.
7.1.2.2 Need to work on the tones

All the students stressed the need for tone work during the first few weeks of starting to learn Mandarin. What constituted this tone work ranged from explaining the differences in tones; getting people to exaggerate the tones and using some kind of visual elements to help people remember tones. The need for visual support for both the tones and the language in general was considered the most important because it aids the memorization process for the language.

7.1.2.3 Reaction to not using romanization

The four students were divided on the issue of using romanization as they were not taught Pinyin from the beginning of the subject. However, being English learners of Mandarin, the urge to romanize was too great. Therefore, from day one of learning, everybody romanized either in English or took notes in Japanese or in Thai. They created their own representation of the pronunciation according to what their ears heard rather than what the Pinyin imposed. The process of creating their own romanization can be an empowering one in that it encourages them to meditate on the language and its pronunciation system. Ideally, students should go through the step of romanizing in their own version of romanization or create their own system of representing the sounds of Mandarin, first with inconsistencies in spelling, then gradually iron out the inconsistencies in subsequent attempts.

In the process, it is possible to enhance students’ powers of perception, encourage them to observe the language and realize that Pinyin is only an artificial system of phonic representation of Mandarin. For example, Student 1 in the control group very quickly realized the artificiality of Pinyin. He stated:
Student 1:

[Pinyin was actually] confusing a bit. …I understood it does not represent the sounds accurately when I first looked up in the dictionary. I couldn”t work out the pronunciation because in English. So, I pronounced it, my Chinese friends say no, it’s nothing like that. Well it’s actually something like this.

However, to recognise the artificiality of the romanization requires students to be sensitive to the linguistic aspects of a language. Student 1 is an experienced learner of FLs having learned Japanese for a long time. So his realisation of the need not to rely on romanization is an exception rather than the rule. The rule is to feel that introduction of Pinyin would made the learning easier. For instance, this was how another student dealt with the issue of romanization:

Student 2:

Last semester I came up with my own Pinyin, which was fine. But it wasn’t particularly standard every time for the same sounds. And looking back to the previous notes of the work, I don’t have the tones noted down. No doubt, this difficulty of having to standardize their own inconsistencies in Pinyin was hard for a number of students.…..And once I have done that [writing in Pinyin] I have got the unfortunate habit which I try to get it out of now of writing in Pinyin before characters…English words.

Pinyin was deemed to be important and essential especially by the less able students in the control group because it was the only tool that linked the sound with tone diacritics. Without any other visual support, tone diacritics, therefore took on much more significance. However, the danger of wholeheartedly supporting the use of Pinyin is that it would lead to students reading and writing only in Pinyin from day one and consequently not learning any characters nor listening to the language at all. Student 2’s comments quoted above testified to the danger of over-reliance on Pinyin.
In terms of getting the students to concentrate on listening and perceiving the language, the design of the course for the control group failed in that students forgot that tones and prosody of Mandarin were just as important as the characters they were learning. All in all, as the interview revealed, the learning experience was not a positive one for most students in the control group.

### 7.2 Experimental group

#### 7.2.1 Results of the computer questionnaires

With the experimental group of students, at the end of the first semester, each student in 2003 filled in the same computer questionnaire and participated in a face-to-face interview with the researcher. In the following sections, specific findings from the questionnaires and interviews are discussed. The discussion included excerpts from the 5 English speakers of Mandarin who participated in 2003 whose data were analysed in detail in the previous chapter as well as other members of the class.

#### 7.2.2 Interview data

##### 7.2.2.1 Direct strategy use of the experimental group

<table>
<thead>
<tr>
<th>Students</th>
<th>Creating mental images</th>
<th>Applying images and sounds</th>
<th>Reviewing well</th>
<th>Employing action</th>
<th>Practicing</th>
<th>Receiving and sending messages</th>
<th>Analyzing and reasoning</th>
<th>Creating structure for input and output</th>
<th>Guessing intelligently</th>
<th>Overcoming limitations in speaking and writing</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>11</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>15</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>9</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>11</td>
</tr>
<tr>
<td>3</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>8</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>11</td>
</tr>
<tr>
<td>4</td>
<td>2</td>
<td>1</td>
<td>0</td>
<td>2</td>
<td>8</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>15</td>
</tr>
<tr>
<td>5</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>6</td>
</tr>
<tr>
<td>total</td>
<td>5</td>
<td>3</td>
<td>1</td>
<td>5</td>
<td>39</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>3</td>
<td>0</td>
<td>57</td>
</tr>
</tbody>
</table>

Table 7.3 shows the “Direct” strategy used by five students in the experimental group.

Compared with the strategies used by the control group, students in the experimental group used
more strategies (a total of 57 as compared to 43 for the control group as shown in Table 7.1). The strategies also varied across the categories.

Table 7.3 also demonstrates that the students in the experimental group used a larger variety of memory enhancing strategies to improve their ability to remember the language. They used “Creating mental images”, “Applying images and sounds” and “Employing action” to enhance their memory.

The design of the course was aimed at getting students to attend to the phonology and phonetics of Mandarin through listening and speaking first rather than reading and writing. This might have been a realistic aim if the learning context was teaching pre-schoolers to speak a FL. However in a tertiary setting, student learning is restricted by institution demands and student styles of learning. In Australia, FL classes are generally organised in 4-5 face to face hours per week over a 13-week semester. Some institutions separate the beginning level subjects into Speaking and Listening streams and streams that focuses on Reading and Writing. At the University of Canberra, the beginning level subjects in Mandarin cover the four macro-skills thus emphasizing reading and writing as well as speaking and listening.

Apart from institutional requirements, adult learners of FLs have had more than 12-13 years of education which emphasised literacy in English. The last time they did any activities in learning to speak and listen in their mother tongue was when they were in kindergarten. Therefore, naturally, when they came to study a FL at university, what they felt most comfortable with was using their reading and writing skills to learn.
7.2.2.1.1 Variety of memory-enhancing strategies used

As predicted, all the students involved tended to start the learning process by working on the characters covered each week on their own. This was perhaps because adult learners are so used to learn through reading, looking up the dictionary and writing characters. Approaching their study this way provides them with a sense of security. However, it can be observed, from the following comments that students in the experimental group used a variety of techniques to try to memorize the characters. They did not just write and rewrite the characters as the students in the control group did. Most students in the experimental group made up flash cards and tested themselves as they either prepared or revised the language learned. Some utilized the audio files to aid their learning:

Student 1:

I have been working on the characters, I make up cards and test myself with cards and go over the weeks materials.

Student 3:

The first thing I will do is I will learn all the characters for that week, I'll do that by making up flashcards and then learning about 5 characters at the time. And just revising them and then using those flash cards to check that I do know them and then build up to 10 characters and 15 characters until I know them all.

Student 4:

Well, I have had to use all the characters from the CDs and then I transfer them all into word documents and enlarge them. That is how I made the flash cards.

Student 4 also applied images to link to sounds. As he commented:

It’s more the audio files on the dataCD because I am using the vocabulary on that and that way I can visualize both of them.
7.2.2.1.2 How did the students practise the language?

Table 7.3 and Table 7.1 show that the number of practicing strategies used by the experimental group of students was 3 times more than the control group’s, 39 and 13 respectively. Apart from the traditional repetition, practicing and writing characters, they made use of the multimedia resources provided. This included the use of the CDs, movies, videos and the Speech processing tool. The following comments describe how students used a range of practising strategies in their learning:

Student 1:

The second half of the semester I have really started to take advantage of the audio CD in my car and the dataCD at home. I am using them more and more now…. The big difference is especially for the language is being able to see something and make the sound for yourself, and then click on the character and have it speak to you and you can check it. I found the audioCD in the car has been great, because that is so much time that has been wasted in the past. And that is good learning time.

Student 2:

Well, I do a mixture of just going through the materials on the CD, listening and making sure that I can pick up what they are saying. The Chinese homework I found is a very good enforced discipline if you like, or set pieces that I found very helpful…..the way the homework is structured is very good and the Taiwanese video clips, I found they are excellent to listen to, I found that they have a good length up to 1 to 2 minutes. To write then the Chinese characters down, I find they are fantastic. And just going through the materials on the CD rom and learning the characters.

Actually, what I just do is, go through the CD, mainly the vocabulary to get used to the sound and get a feel for what the characters look like, and correlate the character and the sound. That is what I try and get a handle on that….. Yes, quite sure, very small talking scenarios. Just buying one or 2 items over the counter sort of thing. So that you get the feel for what you have to do. Get comfortable with a couple of scenarios and then move on.

Student 3:
The first thing I will do is I will learn all the characters for that week, I’ll do that by making up flashcards and then learning about 5 characters at the time. And just revising them and then using those flash cards to check that I do know them and then build up to 10 characters and 15 characters until I know them all. Once I know all the characters and can recognize them, I start putting them to what they actually sound like and say and then go through the week’s work.

Once I know all the characters and can recognize them, I start putting them to what they actually sound like and say and then go through the week’s work and I use the speech analysis tool, and probably would have been using it more, if it had been working in the first part. But certainly, that has been very useful and to help with learning tones and exactly how you do say something.

Student 4:

Well, I have had to use all the characters from the CDs and then I transfer them all into word documents and enlarge them. That is how I made the flash cards. But I use mainly the audio section of the CDs more than anything….It’s more the audio files on the data CD because I am using the vocabulary on that and that way I can visualize both of them. I intend to use the audio CD in the car…… I used the first audio CD quite a bit, I was listening to it all the time in the car, I spend a lot of time in the car….. Mostly, I would get the sound of whatever character I am working at that moment and would write it down about 20 times or so and then just repeat it until I think I have got it. I make flash cards for the different characters and then I try to use the flash card. I use them in 3 stages: trying to recognize the character by sight and then flip it over and work on the tone and the meaning through that way as well as the English meaning……I have had to use all the characters from the CDs and then I transfer them all into word documents and enlarge them. That is how I made the flash cards. But I use mainly the audio section of the CDs, more than anything.

Student 5:

Well, I would probably use the CD-ROM that you supplied but as at home we do not have a computer at the moment that can use it, just mainly when I am studying for the exam, just memorizing them that is just writing them out until you know them, and the vocabulary, just writing it out until you recognize it and for the dialogues, just practicing and practicing it until you can say it. So it is just memorization I think.

From the above quotes from students, it can be seen that students in the experimental group utilized multiple modalities in their learning. They used both the dataCD and audio CD a great
deal; relied on listening both for comprehension purposes and for feedback purposes. Even when learning characters, they created their own flashcards and constantly checked the pronunciation by listening to the corresponding sound files. Most students did not find the task of learning tedious. Their self-learning sessions seemed to be full of variety in all different modalities. For student 5, however, because he did not have a computer at home nor a CD player, he primarily studied by memorizing the characters by writing them out over and over again.

For these students, access to instant feedback was one of the most facilitative aspects of the learning environment under investigation. As the students pointed out:

**Student 1:**

The big difference is especially for the language is being able to see something and make the sound for yourself, and then click on the character and have it speak to you and you can check it. I found the audio CD in the car has been great.

I would like to see more activities and more interaction in class.

**Student 3:**

I would like to have more time on the actual sentences.

The CDs were helpful in learning tones and exactly how you do say something.

[The classes should be] more interactive. I would like to see more activities and more interaction in class, like people in the class do their work before they come to class and I like to have more time on the actual sentences.

Apart from employing more direct strategies for learning, 3 students out of 5 also felt confident in transferring the strategies learned in class to their own learning:

**Student 1:**
[we] got the skills now to sort of take home and to sort of self teach a lot of the characters and the tones.

Student 3:

I go home from Chinese and we have been practicing these things in a group circle and stamping, and I’ll go home and do it home as well: walk around the house and do it, which is why [XX] is also learning. So it affects his life as well. And I remember it through the day and even at work I do it.

Student 4:

I feel a bit more comfortable with it, but I still think that I talk like an Australian. I’d like to be able to speak like a native, but of course that would be pretty much impossible, but I do not feel that I am putting that much intonation into them. I do not feel that the tones are really there. It just feels like if I am saying a word differently as opposed to tones.

From the comment from Student 4 cited above, it seems clear that this student no longer saw speaking with the correct tones as a process of adding tones to characters. Rather, the movement and gesture made it possible for them to “say a word differently” as a whole. Perhaps this student realised that speaking Mandarin involves more than just getting the tones right. In fact it consisted of producing a smooth sequence of sounds with native like rhythm, stress, intonation and tones. As the students in the experimental group practised Mandarin language with its rhythm, stress, intonation and tones in concert in relevant sentences week by week, they experienced first hand, both intellectually and in terms of somatic memories in their muscles, how the different aspects of prosody such as rhythm, stress and intonation, in addition to tones, contributed to one’s ability to speak Mandarin in a way which is more intelligible to a NS in the target culture. Thus the recognition of the importance of saying words differently might suggest that student 4 was beginning to be aware of the importance of the prosodic aspects of speech
involved in speaking Mandarin. No doubt this was the reason why 4 out of 5 students also employed physical actions in their own learning outside class:

Student 1:

I found I was actually remembering how we did all the movements and the intonation, I found I could actually recall them…..I would like to see more activities and more interaction in class and now that we have presumably, got the skills now to sort of take home and to sort of self teach a lot of the characters and the tones, it would be much more valuable to use the in class time more interactively.

Student 2:

Walking around just trying to emulate the sound, hearing and listening, I actually I found them very helpful, for me personally.

Student 3:

Definitely the group work in a circle, clapping, stamping, repeating sentences over and over again is the most useful, more useful than going through the vocabulary.

Student 4:

It sort of made the tones physical. Because you are stamping your feet for 4th tone and sort of raising your body for second tones, things like that. You sort of get an image in terms of physical, what tones represented what.

Student 5:

I do not have anything to compare it with but I do know that if I’d pick up a Lonely Planet guide now, for example and I was reading it to my friend, just the knowledge that I have now on tones, I can make it intelligible for her who is speaking Chinese whereas before, when I was reading it, even if you are reading it and you do not know anything about tones, they do not know what you are saying at all. So I think it has worked.

Student 5 did not participate in the physical movement and gesture phase of the lecture component on a weekly basis. He mainly observed the proceedings on the sideline.
Nevertheless his comments testified that observation of the procedure on a weekly basis was equally beneficial. It helped him to become more intelligible in Mandarin when reading a phrasebook with Pinyin for instance.

Some students were very enthusiastic about including more video skits into the learning environment. However, they were aware that only certain kinds of video materials would be useful.

Student 4:

[I] sort of went through, kind of translate that and pretty much any movie that has got Chinese, I try to translate and see if the translation and the subtitles are the same. Quite often they are not from what I can recognize. But the Emperor and the Assassin, I watched a few times and tried to figure out some parts of that.

Probably when they would have different functions on them with header subtitles, you could have the characters in Pinyin, you could have it in characters, you could have them in English and you could just take them off altogether, so that you could work your way through the stages in the way you felt comfortable. Because like you could hear the Chinese, sort of like following the bouncing ball, they are saying the words, and it would come up in Pinyin, so you could say it. Or you could translate that into Chinese characters, so you could actually see what they are saying.

### 7.2.2.2 Use of indirect strategies by the experimental group

Table 7.4 Indirect strategy use of five students in the experimental group.

<table>
<thead>
<tr>
<th>Students</th>
<th>Centering your learning</th>
<th>Evaluating your learning</th>
<th>Lowering your anxiety</th>
<th>Encouraging yourself</th>
<th>Arranging and planning your learning</th>
<th>Asking questions</th>
<th>Cooperating with others</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>1</td>
<td>0</td>
<td>4</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>0</td>
<td>7</td>
</tr>
<tr>
<td>3</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>4</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>4</td>
</tr>
<tr>
<td>5</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td>total</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>1</td>
<td>4</td>
<td>6</td>
<td>1</td>
<td>22</td>
</tr>
</tbody>
</table>

The students in the experimental group also expanded their use of indirect strategies. Table 7.4 shows that students’ use of indirect strategies spread out across the categories of strategies. This group of students seemed to have been able to organise their own learning much better than the
students in the control group students. One significant strategy use was how they lowered their anxiety.

Throughout the course, the first activity the class did was a relaxation exercise in which students would lie on the floor trying to visualise and imagining themselves being a little white cloud (see Chapter 4, section 4.3.1.1 for a detailed description). Initially students found the relaxation procedures unusual and amusing but most of them participated willingly and found it to be an indispensable part of the classroom procedure subsequently in the course. The reasons behind such measures and their relationship to the process of language learning were discussed with students and students were kept informed of the development of the learning procedure. Past and current experiences of using this technique offered evidence that students reported a sensation of well-being associated with a reduction of failure-anxiety as well as a feeling of alertness that lasts well beyond their classes (Lian, 1980). The following students’ comments echoed Lian’s research findings:

Student 2:

I found actually that in the first half of the semester, for a very challenging language and a new environment, those exercises we did, strangely enough, I found them quite relaxing.

Student 3:

The cloud is helping, because after you work for 8 hours, it is a lot of work to learn Chinese, just to have a break for 5 min. is good just to clear your mind……

Students in the experimental group also spent more time planning and organising their learning than the control group. While only one or two students evaluated and planned their learning in
the control group, in the experimental group, for each indirect strategy (except Cooperating with others), around 3-4 students out of five used every strategy in their learning (see Table 7.4).

7.2.2.3 A student’s reaction to Sptool

Sptool only became fully functional after the 6th week of the first semester, consequently not many students made use of it. However, one student from the experimental group consistently used it and this was how she described her use of the tool:

Student 3:

I listen to the teacher probably about 3 times and then I’ll say it with the teacher 3 or 4 times and then I record myself and look at the graphs and depending on how close they are, which you can get a pretty good idea of whether you are close or not, I will listen to the teacher again and then I listen to my self again just to see how Chinese I am sounding. ..I listen perhaps to the length of a sound and I can see the curve, and if I am missing that when I have recorded myself and I go back and I sort of have a look on the curve where it is quite different, you know where it is a lot longer, then I’ll go back and listen to that exactly character and see whether it is shi or she.

Even though only one student consistently used the Sptool, 3 other students requested more extensive training in using the tool and suggested that the tool should be introduced right from the beginning.

7.2.2.4 Students’ reaction to not using Pinyin

The romanization system for Mandarin-Pinyin- is usually supplied to beginning students in an attempt to make the process of learning Mandarin easier. However, because of the need to increase students’ direct experience of Mandarin in the audition process in this particular experiment, Pinyin and tone diacritics were not provided in the teaching materials. The English speaking Australian students in the course while admitting that the elimination of Pinyin made
learning more difficult, nevertheless felt that in the long run they learned to analyse each character stroke, recognize radicals and associate the sounds of the characters to the characters. Consequently, analysing strokes, recognising radicals and sound association made the learning of the characters easier. 80% of the Australian learners did not see the need to learn Pinyin as demonstrated by the following comments:

Student 3:

I do not like Pinyin. It confuses me. It is certainly not the way I would. I look at how it is and I would not have a clue how to pronounce that and I do not think that now, at this stage it is going to be of that much benefit. And when we have read Pinyin before, it confused me and really it is another language to learn and is there any point in that. If you can learn from the characters, I think that that is probably the best.

That is just another step in the process. And I do not see that there is a real point in that. I mean learning another language, as long as you can see month and you know the character for month and the sound to that, how is knowing the Pinyin really going to help you apart from using a dictionary.

Student 5:

I think that perhaps your emphasis on the tone and not focusing on the Pinyin has helped me with the characters. I think it is painful to go through, to not have Pinyin. But I think that in the end you start to associate the sound with the character which I think is better and also the tones, the focus on the tones is good, because if you do not have tones, you are not speaking Chinese.

Not to learn the Pinyin, because I actually find that if you give me a lot of Pinyin, I couldn’t tell you what it means but if you give me characters I can.

2 out of 5 students acknowledged that Pinyin would be good for semester 2 because it make it easier to look up a dictionary especially when you do not have access to a computer:

Student 5:
But I just found that not having access to the computer, it meant that I would look at a character and I would not know how to pronounce it, that kind of things because I was just looking at the text. But I managed to find what it means. But I guess, I could have saved time if I had a list with the Pinyin, just sort of cheat kind of thing to look at, but not focus on that in classes.

Student 4:

I think that it is handy. It is still a bit early to look at every character and recognize it and put the right tone on it.

7.3 Summary

From both the end of course questionnaires and the face-to-face interviews, it is clear that English speaking students in the experimental group were much more proactive in their learning than the control group. In the process of learning, they were far more self-motivated by creating flashcards for themselves, utilizing a range of learning tools such as the Sptool, audio and dataCDs, video skits, movies, gestures and character sheets to help them learn by themselves. Principle activities included listening to the CDs, linking the sound of vocabulary items to their corresponding characters, using flashcards to revise characters and using video and the NS contacts to test or check their production.

One feature that seemed to generate the most enthusiasm and motivated the learners most was the use of physical movements and gestures in the teaching of Mandarin tones. The physical movements and gestures seemed to have filled the gap for visual reminders of tones that was identified as necessary by the control group students. The exaggerated nature of the movements and gesture certainly created very strong memory traces in the students that resulted in students dealing with the issue of tones and prosody of Mandarin in various ways. The use of the physical movements in this study was motivating because
It was an active process because I had to ask questions and be able to ask questions…. because the movement made me more motivated to learn more Chinese because after doing that I felt good, I felt confident, I felt that actually I had actually done something “ (Student excerpt from Interviews from the experimental group in 2004).

All the students from the experimental group felt confident after doing the physical movements. This contrasted very well with their attitude toward the more traditional style of teaching:

I think there is something in the process of action that makes you feel that you have achieved something. And I think that the passive way of learning of just sitting at a seat and having a teacher just to tell you that this is the thing, this is the thing etc. removes you from the working out process that needs to take place within your brain or yourself (Student excerpt from Interviews from the experimental group in 2004).

Furthermore, the end of semester questionnaire shows that when comparing the amount of time on task by the experimental and control groups of students, experimental group spent 3 times longer (10 hours per week versus 3 hours per week) than the control groups of students in their self-study even though 50% of the students in the experimental group were mature age students who worked full-time. The students in the experimental group had an average age of 30.25 years and were highly motivated to learn Mandarin. The average age of the control group was 24 years old. Therefore, in this study, the experimental group of students seemed to be so motivated by the design of the learning environment that the influence of many inhibitors, such as age, postulated by the CP hypothesis, was, to a certain extent, reduced.

In conclusion, students in the experimental group seem to see that learning Mandarin using SEA was a positive experience. This contrasts sharply with the difficulties that the control group experienced with the conventional communicative methodology.

The conclusions of the study will be discussed in the next chapter.
Chapter 8 Conclusions

This thesis has endeavored to evaluate the efficacy of SEA for language learning for a group of Australian students. The insights from diverse fields such as cognitive psychology, L1 acquisition, SLA, neurology, biology, psychotherapy and the field of corrective phonetics have all contributed to the creation of such an approach. The design is not meant to be “all-inclusive” as such it cannot nor should be considered as ever being completed, definitive or final. Instead, it is hoped that the role this thesis has played is not only to provide “knowledge”, but to initiate a discussions about learning and research in SLA in a more adventurous way.

In evaluating the effectiveness of SEA for the students, several facts, constituting a body of knowledge, have been collected with regards to the students’ experiences of learning Mandarin as a L2. In the first part of this chapter the major findings of the present study are summarized. In the second part, issues and principles that emerge from the study will be considered and discussed and lastly, future lines of research are explored.

8.1 Information obtained about the process of prosody acquisition by English speaking Australian students learning Mandarin as a second language

In investigating the process of prosody acquisition of Mandarin by a group of English speaking Australian students learning Mandarin as a L2, the following information was sought:

1. Sequence of mastery of the four tones, and hence the order of difficulty of those tones;
2. Common tone error patterns among learners;
3. Likely cause(s) leading to the detected errors;
4. How and where interference from English intonation takes place and how significant is its role;
5. Effectiveness of SEA on the teaching of Mandarin tones and prosody;
6. Effectiveness of specific computer enhanced learning programs in the learning process.

Results relating to items 1 to 6 were described in detail in Chapter 6 and 7 of this thesis.

Conclusions drawn from the results will be summarized below.

8.1.1 Order of difficulty with the four Mandarin tones

As reviewed in Chapter 3, section 3.2.1.4, researchers in the field of teaching Mandarin as a L2 could not reach a consensus on the order of difficulty of tones in the learning process by American learners of Mandarin (Chen, 1997; Miracle, 1989; Shen, 1989b). This study reveals two interesting phenomena:

1. Even though two separate groups of students were involved, there was a similar system of errors made by these students. This supports Lian’s argument that when learning a L2, students from the same L1 background tend to generate a similar system of errors (Lian, 1980);

2. The kind of input and testing mechanism used to elicit data might have affected the order of difficulty of tones obtained.

Unlike the designs of Miracle, Shen and Chen’s experiments, the experimental design used in this study tested the use of language in situations closest to how the students would use the language in real life. The resulting orders of difficulty obtained from both control and experimental groups seemed to support the view put forward by Clumeck (1980) that the frequency of certain prosodic cues in the input is more likely to be instrumental for the acquisition of a tone than any particular innate degree of difficulty. In other words, in the learning environment under study here, the teaching materials used (which constituted most of
the input taught to the students) and the teacher directed methodology both played decisive roles in determining the order of the tones acquired.

Ellis (1999:161) suggested the following features are likely to facilitate rapid development in SLA:

1. A high quantity of input directed at the learner.
2. The learner’s perceived need to communicate in the L2.
3. Independent control of the propositional content by the learner (e.g. control over topic choice).
5. The performance of a range of speech acts by both NS/teacher and the learner (i.e. the learner needs the opportunity to listen to and to produce language used to perform different language functions).
6. Exposure to a high quantity of directives.
7. Exposure to a high quantity of ‘extending’ utterances (e.g. requests for clarification and confirmation, paraphrases and expansions).
8. Opportunities for uninhibited ‘practice’ (which may provide opportunities to experiment using ‘new’ forms).

The pedagogical design of this environment seems to have exerted a powerful influence on how Mandarin is learned. SEA contains all the features mentioned above with possibly items 6 and 8 lacking as in a FL learning situation as opportunities for uninhibited practice are not always readily available. However, SEA offered many opportunities for students to have direct contact and exposure with the TL thus resulting in L2 students being able to speak intelligible Mandarin after just 6 weeks of learning. The opportunities for the direct experience of the TL came in the form of a teacher-led classroom procedure, easy access to the TL through CD-ROMs (see Chapter 7, section 7.2.2.1.2) and a curriculum relevant to the needs of the students. In other words, even though students were not taught according to some “natural” order of tone acquisition, L2 students still achieved the goal of being able to speak intelligible Mandarin.
The findings of this thesis, in short, are that English speaking L2 speakers of Mandarin learn tones in a FL context in no "natural” order. Furthermore, there is no evidence that one tone is more or less difficult than any other; it might all depend on what input they have been exposed to and how they have been taught to deal with the Mandarin input. In SEA, in the space of six weeks, some tones were learned more successfully than others. Even the more difficult tones such as Tone 3 are partially under control for some students (see page 181 on the discussion of Student 1’s performance in the experimental group).

8.1.2 Common tone error patterns and causes among learners

The kinds of errors produced by the two groups of students were also similar. These errors were:

1. The beginnings and ends of sentences presented most problems for both groups of students.

2. When producing a sequence of items, both groups of students had a tendency to produce them as a series of rising tones until the last items. This contrasted clearly with NSs’ productions of the same sentences.

3. Specific errors with the four tones.

The two groups of students made similar errors with the four tones. For a detailed description of tonal errors, please refer to Chapter 6, section 6.5.4. The most problematic tone seems to be Tone 3 as in running speech, all the allotonic variations of Tone 3 are hard to predict. As Shen points out that Tone 3 changes “according to the phonological rules” because “the allotonic production of Tone 3 depends idiosyncratically on speech tempo and stress and it varies from utterance to utterance, from speaker to speaker even within the identical linguistic context”. Therefore, it is not surprising that so many Tone 3 errors were also made by students involved
in this study. The large number of errors in Tone 3 in this study confirms the conclusion reached by other researchers that Tone 3 is difficult to acquire due to its complex tone sandhi rules.

Realistically, outside the classroom, one seldom listens to or produces isolated syllables. For students to acquire the correct tones in connected speech, it is necessary to treat tones resulting from tone sandhi as the norm. In this case, students should be taught through targeted exposure to many realizations of Tone 3 in the context of connected speech. Students in this study were not taught Tone 3 with any special treatment except through normal exposure of the teaching materials. However, due to students prolong exposure to the CDs used in this course, there are signs that even Tone 3 is partially under control for some students. Please see section 6.4.3, Chapter 6 for a detailed discussion.

With respect to Tone 4, the production of both control and experimental groups of students was influenced a great deal by the intonation pattern of English that is their mother tongue. First, some students did not start Tone 4 at a high enough point in the tone space. Second, though all students could produce a downward contour, students did not go low enough with Tone 4. This inability to reach the lowest level of their voice range was also compounded by the large number of Tone 4 present in the input. In the teaching materials and oral performances collected, the proportion of Tone 4 was the largest out of all the other tones. Furthermore, many Tone 4s occurred in the second last syllables of questions forms involving the use of question particles such as ‘ma’ or ‘le’. Consequently, many Tone 4 words were produced with an uptrend at the end of a question as some students automatically transferred English language intonation pattern for questions when producing Mandarin questions.
8.1.3 Likely causes for the detected errors

Many researchers have theorised the likely causes for the errors in Mandarin prosody especially tones. However, there is no agreement on the order of difficulty of tones being one of the likely causes for errors in this area. Another likely cause which has gained agreement from all researchers is that L1 interference from L1 is likely to be the primary cause of many of the errors produced by students. The present study concurs with this view. SEA has achieved limited success in combatting the interference from students’ L1 which is English. The specific practices used in SEA are:

1. Presenting the learning materials in written form in Chinese characters rather than romanization.

An alphabetic language such as English uses rules that do not coincide with Pinyin. Thus using Pinyin would only activate the transfer of English linguistic habits to Mandarin. In short, if L1 interference were considered the primary barrier to communicative competence in learning Mandarin, then the representation of Mandarin by the Roman alphabet can only be construed as an enhancement of this difficulty. One of the biggest problems facing adult learners of Mandarin is the need to use their memory to learn the language. Learning Mandarin particularly relies on memory because the written language is not phonetic. In other words, the written language cannot be used as a tool to remind students the pronunciation of the words without learning how to read the words first. Over-reliance on Pinyin is dangerous if Pinyin is used because it causes students not to listen to the language nor memorise the language or memorise any characters. The experimental group of students overwhelmingly rejected the use of Pinyin because they felt that Pinyin is another artificial system of representation of the language.
Learning it is only likely to be confusing. Students found it better to listen to the language rather than learning an artificial system of representation.

Anecdotal evidence suggests that students taught with SEA (experimental group) might not know how to spell the language in Pinyin nor knowing which tone is attached to which character but when they speak Mandarin what reminds them is the rhythm and melody of a particular sentence. Usually what triggered the memory was only the beginning of a rhythmic or melodic pattern. This finding seems to fit in with the multi-trace theory of memory proposed by Goldinger (1997). Students further report that when speaking Mandarin other triggers seemed to operate to remind them of the language learned. These triggers were visual image of a page; visual image of a word on a page; or the sound of a word in a sentence.

2. Through the provision of CD-ROMs, students’ opportunities to experience the TL, as suggested by Verbotonalists (Renard, 1975), increased enormously as shown by the experimental group. They did not seem to have been disadvantaged by the removal of the traditional “helpful” tools such as tone diacritics. The fact that they were able to receive a large amount of direct experience with the TL enabled them to comprehend and be understood by NSs. This further bolstered their confidence and motivation in using the L2.

3. Feedback procedure using movement and prosody rather than lexical tones proved to be more effective for students’ learning in the experimental group. Students in the experimental group were so motivated by the body movement procedure that they expressed the need to spend more time on this procedure. Intellectualising about Mandarin (in terms of explaining grammar and about Mandarin language and
translation) is kept to a minimum in the two-hour sensitization session because it does not contribute to the direct experience of TL.

4. With prosodic features such as intonation patterns, stress, and intensity preserved and taught in relevant contexts, students not only learned the correct tones they also learned rules about stress, pauses, intonation and grammar *implicitly*. As stress in Mandarin is also tied to grammar, learning utterances, therefore, also means learning the grammar at the same time. Research in L1 acquisition has found some evidence to confirm that prosodic bootstrapping occurs in infants acquiring L1. It seems that it is highly likely that prosodic bootstrapping can also happen with adult learners of a L2. It might be that prosodic bootstrapping is responsible for the experimental groups’ longer and creative language use as well as increased intelligibility of the Mandarin they produced as judged by a group of 9 independent Mandarin NS judges. It can be argued that students taught with SEA perhaps have developed a better “feel” for the Mandarin language which enabled them to produce more acceptable spoken language in Mandarin. However, what exactly this “feel” is or whether it has any concrete existence is a matter for speculation.

5. Another influencing factor could be that students had easy, on-demand access to learning materials in audio or written format. This made the learning easier and handed over control to students to determine how they wanted to interact with the material.

6. Another essential characteristic of SEA is the affective learning atmosphere created by the use of relaxation technique and the walk-in-a-circle movement procedure. Like the Verbotonal method, SEA does not see capacities possessed by students through their L1 in a totally negative light. While the knowledge of their L1 phonological system might indeed act as a “phonological sieve” as proposed by Trubetzkoy (1939), the reactivation
of other senses and mechanisms in the learning process such as the “soma” produced highly motivated and socially confident students. In Dijohnson and Craig’s study (1971) it was found that deaf children learning under the Verbotonal method achieved significantly better gains in lipreading and social competency. Dijohnson concluded that “perhaps the reciprocal nature of these competencies—lipreading and social skills—with that of speech production is indicated” (DiJohnson & Craig, 1971:31). The results of this study seem to support Dijohnson’s conclusion. In other words, learning with SEA appears to offer a setting in which previously acquired social skills of adults can be effectively utilized. SEA does not see adult L2 learners’ past history as hampering the learning of a L2.

7. The consistent practice of various learning strategies such as using relaxation to create a positive learning environment, body movement, gesture, humming, mouthing to the words and so on taught experimental group of students not only to use these strategies in class but also transfer them out of the classroom into self study situations. This suggests that it might not be at all necessary to come up with a precise constellation of learner characteristics which good learners engage in to guide future generation of language students. A language learning environment can be designed along the principles of SEA to utilize what is already within students’ capability for language learning.

In short, these practices have been effective for students taught with SEA. The ultimate indicator of its efficacy comes from the fact that students in the experimental group significantly outperformed the control group in the intelligibility test by 9 NS judges.

SEA particularly benefited male students in that their voice range when speaking Mandarin changed dramatically. Acoustically, they spoke Mandarin with a much wider fundamental
frequency ($F_0$) distribution (measured acoustically in group Mean $F_0$ SD) than when they spoke English. This perhaps constitutes the first requirement for attaining native level Mandarin according to past research conducted by Chen (1974).

The difference in the voice range of students found in this study suggests a different direction in the acquisition of Mandarin prosody should be entertained. Instead of concentrating on a search for the developmental order of Mandarin tones, research into how students can be made to expand their voice range and how such students might develop in their future study might be warranted. It might be reasonable to assume that in order to successfully “teach” Mandarin prosody, the first task is to teach students ways of expanding their voice range when speaking Mandarin.

8.2 Discussion of principles

8.2.1 Contribution of gesture and movement to the teaching of Mandarin prosody

One of the useful tools acknowledged by the experimental group of students was the extensive use of movement and gestures. The exaggerated nature of the movements and gesture certainly created very strong memory traces in students thus enabling them to deal with the issue of tones and prosody in various ways. So, why does it work?

The traditional method of teaching L2 tends to focus on explicitly teaching grammar or teaching students about language. In other words, Mandarin is taught usually by treating it as a separate object to by the student. As such, this object possesses predefined signification, denotations and connotations. Because of this students try to observe the behaviour of the foreign object, learn the denotations and connotations of the object and then copy them and get the foreign object into their heads. The process is usually "conscious" and "rational". This is not likely to work if we aim at producing students with highly intelligible speech in Mandarin because our L2
learners would have practised and perfected the “form and structure of the language system of his culture…[which] would encompass a multiplicity of interlocking aspects: rhythmic and syntactic “hierarchies,” suprasegmental features and paralinguistic nuances, not to mention body motion styles and rhythms” (Condon & Sanders, 1974) since the first day of their lives.

In SEA, by working on sentences and rhythm through student-generated movement and gestures, students physically realign their proprioception through the soma, the skin, the sensation they feel when they move and in the pit of their stomach (Lian, 1980) as well as the vibration they feel in their ears and in their bones (Renard, 1975). It is postulated that in this way L2 students can expand the entrenched rhythmic and syntactic “hierarchies” to include Mandarin rhythms and syntax. Therefore, SEA does not merely utilise student's powers of cognition but finds a way of “entering” the prosody of the TL through the listening and practicing of rhythms. Students actually learn by being “in” the language in a new sense. They are cathected by the TL and impelled by its rhythms. In other words, they discover the TL as though it were within their body rather than encountering the language as an object "outside" themselves (as in the traditional pedagogy).

The state of mind students get into is perfectly described in the following quote:

> It was an active process because I had to ask questions and be able to ask questions…. because the movement made me more motivated to learn more Chinese because after doing that I felt good, I felt confident, I felt that actually I had actually done something. And I think that the passive way of learning of just sitting at a seat and having a teacher just to tell you that this is the thing, this is the thing etc. removes you from the working out process that needs to take place within your brain or yourself “ (Student excerpt from Interviews from the experimental group, 2004).

In this sense, while they are listening to the language, they are also producing the language in the next breath and be understood immediately. In this sense, the mastery of the language is
achieved actively, experientially and holistically. Students actually feel that they are achieving every time they listen, move and produce. In this sense, the learning process is existential and enunciative (Deleuze & Guattari, 1987) rather than cognitive. The different steps of humming, clapping, using gestures, mouthing and finally speaking the sentence normally not only allow learners to experiment with different ways of perceiving the same language, they also allow them to gradually build up confidence in uttering the sentences more and more accurately.

8.2.2 Language learning needs to be carried out in its social context and access to the TL discourse community is important

The debate on the relative importance of learning styles and the socio-cultural environment of learning is far from over. It is probably safe to acknowledge that both are important for learning a L2. This study’s contribution to this debate is that it shows that a motivating learning environment can be constructed to cause both individual and socio-cultural strategies to be activated. The relaxation and the walk-in-a-circle procedures appears to have produced, consistently, a relaxed social atmosphere and a sense of well being in students while in class conversation activities cause a community of learning to be established among students. The positive transfer effect of the classroom procedure is also evident in the efforts students in the experimental group made to maintain their contact with NSs outside class even though many of them worked full time and are mature students. Having access to the TL discourse community is extremely important because it allows students to link the language learned to the world of experience. As Oller argues:

The learner must perform the pragmatic linkage of meaningful texts in the TL to his or her own experiences. Without access to well-equilibrated pragmatic mappings of target-language texts, language acquisition cannot occur.
Of equal importance to the kind of data provided is the learner’s perception of himself or herself as really needing to understand and make use of the TL to such an extent that he or she in fact does so. Saying that one is motivated to acquire a language means nothing if it does not motivate active participation in the process of pragmatically linking texts in the TL with the learner’s own personal experience. Merely being exposed to pragmatic mappings of target-language texts into someone else’s experience will not be sufficient. The texts must be actively and successfully mapped into the learner’s own experience. (Oller, 1991:4)

This study shows that, because of SEA, the experimental group of students vocalised more by producing longer and more complex utterances (see Chapter 6, section 6.5); were extremely motivated as shown by their involvement with the Mandarin speaking communities and the range of learning strategies used (see Chapter 7, section 7.2) and did not seem to be hampered by age- the primary factor in adult learner’s poor performance as proposed by supporters of CP for SLA.

8.2.3 The role of CD-ROMs

The experimental group of students spent a considerably longer time on task and used the various computer enhanced learning tools in a variety of ways. The most popular tool was the audio CD as it required only a CD-player. No doubt this easy access to the TL further increased students’ direct experience of the TL. It must be borne in mind, however, that whatever technology one chooses to use to enhance learning, they must be designed according to pedagogical principles. For instance, the technology used in this study (CD-ROMs and Sptool) were designed to provide students with increased opportunities for direct experience of the language and for exploration.

8.3 Implications for course design

The selection of course materials for SEA is crucial. In selecting teaching materials, the following principles should be adhered to:
1. Principle of catering to students’ urgencies

The selection of the teaching materials should be based on the urgencies of the students. For instance, in this study, students felt that they were learning things that they can use and hear in real life. Of course, in this case the best source of learning material is then authentic materials.

2. Ideally sentences should be presented with their prosodic aspects preserved and if possible as close as possible to the form used in authentic communicative situations.

The most important factor in the success of SEA is to teach sentences (not words) with their prosodic aspects preserved. The ramifications of teaching sentences are varied. Learning sentences with their stress, rhythm and intonation intact not only aids memory (Mandel et al., 1994; Mandel et al 1996), it might also help in the acquisition of Mandarin syntax according to research findings from L1 acquisition. Learning sentences will also achieve better results in communication.

In FL situations, however, sometimes it is difficult to obtain authentic materials. This problem can be solved if students are asked to test what they have learned by listening to authentic interactions sourced from the TL.

3. Prioritising speaking and listening rather than reading and writing

The written Chinese language is composed of ideographic characters. When one sees a word or a sentence in alphabetic languages such as English, there is a relative transparent relationship between spelling and pronunciation. When one sees a Chinese character, however, it is not pronounceable unless one has learned to pronounce it. Furthermore, when a line of Chinese text is written, there is no physical space between each character or each word announcing
physically the separation of words. This is further complicated by the very definition of a “word”. For instance, in English, the definition of a word is easy. A reader only has to identify the spaces between words to see a word. But how would a reader know where a word starts and where a word ends in Chinese when the characters are not divided by spaces?

In this sense, learning sentences with the respective prosodic markers intact also help to sensitise students to a whole host of linguistic phenomena in Mandarin such as morphology and syntax without having to resort to a metalanguage to describe grammar or morphology. This probably accounts for the lack of problems with pauses in the data of the experimental group of students. Consequently, teaching with SEA could mean teachers would be conveying information about Mandarin morphology and syntax and a whole host of other areas without having to use metalanguage to describe, for example, grammar.

4. Teaching L2 by allowing learners to “enter” the language through body movement and gesture.

By allowing learners to learn through body movement and gesture, we are actually designing to change the way they carry themselves, their proprioception and the way they interact with the world at large. In the process, it is irrelevant whether students have managed to get rid of “sieves” (be it phonological, cultural, or philosophical); what is relevant is that their L1 rhythmic and syntactic “hierarchies,” suprasegmental features, paralinguistic nuances, body motion styles and ways of seeing themselves have been modified. Culturally, this contributes to the principle of openness that encourages students to take into account the ontological complexity of culture and language rather than trying to reduce the world according to a set of theories, perspectives, approaches and other variables.
5. The principle of the social individual

Despite the social construction of an individual, the unique history and perspective of the individual should be catered for in the learning process. SEA is constructed according to this principle with its inclusion of technological tools to cater for individual learning needs and procedures for building an inclusive and relaxed learning atmosphere.

6. The principle of modesty.

Kraus (2003:361) suggested that “the modesty principle needs to be adopted so as to make pedagogues accept that they do not nor will not have all the answers” in designing learning environments in language learning. In fact, SEA is no more than a suggestion for changing the pedagogic practice in the field of teaching Mandarin as a L2.

8.4 Contribution of this thesis to the state of knowledge

The findings of this research are wide ranging and likely to be of value in that:

1. It is the first study in the field of Mandarin language teaching that brings the object of study into the proprioceptive/psychological system of the learners rather than keeping it separate from the students’ psychological systems.

2. This is also one of the few studies to recognise the relationships between language, gesture, movement and rhythm and to use the understanding of such relationships to create an effective learning environment.

3. It outlines the implementation of a learning methodology in which body movement was central to learning. Although this research was not carried out in therapeutic contexts such as Dijohnson (1971) and Brüll’s (2003) studies, the findings of this study provide support to the findings from Dijohnson (1971) and Brüll’s (2003) studies. This study
provides strong evidence that the use of body movement in language learning can be extremely effective.

4. Apart from the use of body movements, the procedure of walking and clapping to the rhythm of the language was extremely important in potentially setting up a conducive state for learning. In teaching students to assimilate the body’s motion and senses and to apply them to academic learning, the author of “Somatic Education for children”, Susan Kramer (1998-2005) often uses clapping with the rhythm of the language to increase the power of concentration and acuity in hearing. Clapping is also used to make the learners feel the beat within the body in her work. Similarly, the students in the experimental group and other students using SEA in subsequent cohorts reported the effect of the classroom procedure to last for a number of days. They reported that it was during this state that they did revision work themselves. This could be one of the important reasons why the students in the experimental group spent much more time engaging with the learning tasks in self-learning. Thus this study provides evidence that SEA contributed positively to many aspects of L2 language learning even though many effects of SEA are not describable or knowable.

5. The present research also represents a study that attempted ways of creating successful strategies for learners through the enrichment of a range of available tools and resources. Many learning strategies used by good language learners were adopted by the experimental group of students. This suggests that it is possible to change the learning behaviour of students if the learning strategies are embedded in the environment in which they learn.

6. It is also the first study that investigates the pitch range of English learners speaking Mandarin.

7. In terms of teaching of Mandarin phonology, a learning environment constructed according to the principles of SEA not only can implement a number of suggestions put forward by previous researchers but also illustrates how other neglected areas of teaching such as stress and intonation can also be taught in the same environment.
8.5 Unresolved questions

8.5.1 Role of romanization in Mandarin language teaching

The question of if and when to use Pinyin romanization still remains unresolved. Most students acknowledged that romanization was not helpful in the initial stage (from the first day of learning till the end of first semester) of learning Mandarin. The new tools, on the other hand, were extremely well-received by the experimental group of students because they enabled students to receive the audio-stimulation of the TL in a way similar to that which L2 learners would have to deal with in real life. In other words, they had to deal with the new language right from the beginning. Thus, in the initial stage of teaching Mandarin without Pinyin and diacritics is not only possible, it may even be preferable.

However, students in both groups acknowledged the usefulness of Pinyin after the initial period. It is essential for the students not to become over-reliant on Pinyin which is, after all, printed spelling of the sounds of Mandarin. The question of when romanization should be introduced needs further study. However, to conduct the research, it is important to move away from the research methodology of analysing data obtained from a class of students using romanization without examining the interplay of other factors in the learning environment. As the instruments used to collect data can determine the research findings obtained, in research on the role of romanization in teaching Mandarin, it is important to make sure that the instruments used to collect data (which is used as an indicator of student L2 attainment) do not involve the use of romanization. If romanization is only a means to an end which is the mastery of Mandarin language, then what we need to know is whether and how the learning process (with or without using Pinyin) would benefit students’ ultimate attainment in Mandarin.
8.5.2 The use of Sptool

Though the benefit of Sptool was recognized, only one student consistently used it and three others expressed the wish to do so. The mitigating circumstance against the widespread use of Sptool was basically a technical one because many students experienced compatibility problems in installing the tool. Thus the findings of this study do not contain enough evidence to confirm Hardison’s (2004) findings that training with computer-assisted visual feedback tool can produce a significant improvement in both levels of the spoken language with evidence of generalization in novel stimuli. Specialised computer enhanced learning activities targeting problem areas such as the perception and production of Tone 3 and 4 have the potential to further lessen the influence of L1 interference and overcome the inherent difficulties posed by Mandarin.

Obviously further research in this area will be helpful in understanding the role of such technology in learning. However, one needs to be cautioned about conducting such research in the use of Sptool using a conventional teaching methodology. As described in Chapter 4, the development of Sptool was directly related to the development of SEA. Thus, an investigation of Sptool should also be conducted in a learning environment where SEA is being used.

SEA is also part of a much wider language learning environment which can easily incorporate other computer or non-computer learning (not just language) techniques. Other possible computer enhanced learning tools that can be incorporated into this environment include Mmbrowse (Lian, Lian, & Zhang, 1997b), MMBase (Audiovisual database) (Lian & Lian, 1997), Intonation Patterns of Chinese (Lian, Lian, & Zhang, 1997a), WWTLab (a dialogue practice system)(Lian & Lian, 1996). The philosophy and flexibility of this learning environment makes it very much a user-based environment.
8.5.3 Is this methodology similar to that of the Total Physical Response (TPR)?

The question might arise as to whether this methodology is similar to the Total Physical Response (TPR) (Asher, 1977) methodology developed by Professor James Asher in 1977. SEA is not like TPR because:

1. It does not adopt the strategy of delayed speech in the arrangement of learning activities and materials. In fact, students are required to produce almost immediately.
2. It acknowledges the differences between L1 and SLA in that it promotes the motivational energy engendered by adult students’ social needs to communicate.
3. It does not organise teaching materials and activities on some arbitrary principle of difficulty or developmental sequence. The materials were chosen first and foremost according to the principles of communicativity and relevance to the students.
4. Apart from teaching the language, students were also taught a number of processing load lightening skills so that they learnt how to physically, not just mentally, produce the language in its optimal prosodic contexts. The physical gestures for tones also equip them with physical reminders that tone is an indispensable part of Mandarin pronunciation. In other words, the approach does not work on one isolated aspect or element of Mandarin. It works simultaneously on segmental, suprasegmental, grammatical and syntactic systems.
5. Most importantly SEA does not use gesture as a mnemonic devise for enhancing the acquisition of vocabulary items only. It acts as a reminder of a whole set of known and unknown memory traces. SEA may also use movement and prosody to help students to segment the language stream.

Asher’s approach to language learning, however, assumes that listening should be developed before speaking because this is what happens when a child learns his/her L1 and delaying speech seems to reduce stress (Gary, 1975). In adults who are learning Mandarin as a L2, obviously the proposed constructs of phonological store and articulatory rehearsal (Gathercole & Baddeley, 1993) might already be fully functioning in their L1. Therefore theoretically, even
if we want to delay production in L2, the developed coordination in the phonological store and
the articulatory rehearsal might not be able to be stopped.

Secondly, as language learning theories have convincingly argued, the need to keep the
language learning process communicative for social reasons, it is both theoretically unsound
and potentially de-motivating if adult learners are prevented from trying out the newly learned
language through oral communication.

In terms of classroom activities, Richards and Rodgers (1986) state that the labelling and
ordering of classroom activities in TPR seem to build on the structural view of language in that
the mastery of a language is through the sequential mastery of phonemes, grammatical units,
lexical items and grammatical operations. One key objective of TPR is the use of action-based
drills in the imperative form. It is difficult to see how these commands are linked to the L2 as a
whole apart from being an element of it. SEA, on the other hand, does not order classroom
activities according to structural views. It is based on the communicative needs of students.

8.5.4 Does metalanguage aid learning?

In the course of conducting this research, one question concerning the construction of the
learning environment has emerged. In assisting students’ engagement with the target culture,
are devices such as Sptool, body movement and gesture, CDs, grammar points, metalanguage
for discussing grammar and/or phonetics, tone diacritics, romanization and translation equally
useful to students?

In a student’s entire Mandarin learning career, all these devices will be useful at one time or
another. However, what this research has shown is that it is very likely that at different stages of
learning a L2, some devices will be better for achieving the goal of learning than others. In the
case of this study, the primary goal of total beginners of Mandarin is to acquire the phonetic and phonological systems of Mandarin through audition. Audition inevitably involves a large amount of listening and attending to the segmental and suprasegmental structure of Mandarin. The finding of this research shows that the use of body movement and gesture, CDs combined with not using tone diacritics, romanization, grammar discussion, metalanguage to talk about the language and translation have enhanced students’ ability to listen to Mandarin input and successfully acquire the segmental and suprasegmental structure of Mandarin. However, this is not to say devices not employed in this implementation of SEA are not going to be useful at other stages of their learning. For instance, when the goal is for learners to acquire Mandarin syntax, due to the non-inflectional nature of Mandarin, students not only need to attend to the relevant grammatical particle or point in isolation (i.e. in academic discussions in class), they also need to attend (notice, or hear) that particular grammar item in the running speech of NSs. Metalanguage is then extremely useful in academic discussions as aids to understanding.

Whether it helps a student in real life is not clear. Students also need to practice listening to NS speech in order to practice picking up the relevant grammar point in context as most of the grammatical items in Mandarin tend to be unstressed. This difficulty of picking up grammatical items is further compounded by the lack of redundancy in Mandarin. Ultimately, in real life, it is up to the individual student to pick up the relevant grammatical cues in the NS’s language and respond appropriately. No teacher on earth can force students to attend to a particular item in real life discourse. What we can do, however, is to create a conducive environment in which students practice with materials which are as authentic as possible so that they can test out their hypothesis about the language in a safe environment.
8.6 Future research directions
8.6.1 Some new questions to be asked
8.6.1.1 Do L2 learners of Mandarin segment Mandarin as L1 children do?

The research undertaken in this thesis is fundamentally different from previous research in L1 and SLA of Mandarin (Clumeck, 1977; Hsieh, 1996; Li & Thompson, 1977; Lin, 1985; Miracle, 1989; Shen, 1989). Instead of assuming that input is the same for every L1 infant, thus legitimising a search for developmental sequence, this thesis questions the way infants deal with Mandarin input. How do they segment the input? How do they develop sensitivity to certain characteristics of Mandarin such as tones? Research findings from L1 acquisition research of other languages offered some partial answers to these complex questions.

This thesis is an investigation of a teaching method for L2 adult learners of Mandarin. Insights from L1 acquisition were helpful in focusing this research on how input should be provided and what measures can be built into the learning environment to make the job of segmenting Mandarin input and developing sensitivity to characteristics of Mandarin easier.

An alternative starting point of L2 teaching pedagogy and research could be on suprasegmentals rather than on segmentals. In research on the teaching of suprasegmentals, different questions might be asked. Questions such as: What do adult L2 learners do when segmenting a language stream? What constitutes quality input for learners? How do L2 learners listen to a string of L2 input?

8.6.1.2 Does this study support the critical age theory of language acquisition?

performance in a L2 by later learners might not be attributable to the reduced plasticity of the brain or the fossilization of L1 linguistic habits. An alternative explanation could be that older students’ needs or priority in language learning might be different. For instance, for an adult living in say the People’s Republic of China, the communicative need might be to solve a series of practical problems such as renting a house or opening a bank account. In performing such tasks, the communicative need for the L2 adult learner is likely to require better linguistic competence than is required of a L2 child. For such adult learners, they might think that they do not need to sound like NSs as long as they succeed in performing the necessary tasks.

Consequently, in learning Mandarin, their attention might be directed to other aspects of the language such as syntax. L2 children’s primary need is likely to be concerned with the need to socialise and get on with their Chinese speaking friends in schools so the need to sound native is paramount. Furthermore, linguists have often perpetuated the myth that after puberty, a L2 learner is unlikely to achieve NS pronunciation. It is little wonder that many adult learners do not even contemplate the possibility of achieving near native pronunciation in any L2.

In this study, the fact that students in the experimental group were older and yet outperformed the younger students in the control group provides some evidence that age might not be the only crucial factor affecting the acquisition of pronunciation. It is possible to argue that the conditions in which learning take place might be just as important. This study was conducted in a learning Mandarin as a FL context. In other words, the input available to students was mainly through the face-to-face contact hours and through their course CDs. Consequently, the input available was restricted. This restricted input, in fact, can be said to have facilitated learning as students could be directed to focus on certain learning areas such as audition. Learners did not have to cope with tasks that required high level communicative competence. This enabled them
to make use of opportunities that allowed them to engage in “sound play” with peers thus
gaining a better command of the phonology as suggested by Peck (1978).

8.6.1.3 Is there any evidence to support Krashen’s “input hypothesis”?  

In SLA, many theories have been proposed to explain the language learning behaviour of L2
students. Since the proposed “input hypothesis” by Krashen (1985), many researchers such as
Swain (1985) have argued that output is equally important in SLA. This thesis offers some
evidence supporting Swain’s “comprehensible output hypothesis” but points to the difficulty of
a wholesale adoption of Krashen’s “input hypothesis”. It can be seen from discussion on the
order of difficulty of tones in Chapter 6, section 6.4.3 that because of the difficulty of defining
and measuring variables such as prior exposure to a L2, both “i” (input) and “1” in Krashen’s
“i+1” can be no more than metaphors to guide the sequencing and selection of learning
materials.

Textbooks and curriculum in L2 teaching of Mandarin have been designed according to the
“easy” (e.g. one syllable”) to “difficult” (e.g. two syllables”) formula for years with apparently
limited success. This might be because what is considered “easy” or “difficult” by teachers
might not be considered in the same light by learners because of their diverse learning needs. In
the present thesis, the selection of learning material is based on such learning needs. In other
words, SEA creates an environment that enables learners to make choices about what they wish
to focus on. It also allows learners to develop personal system for remembering many aspects of
Mandarin that they need to master. As a result, SEA enabled learners to produce better
Mandarin. Quantitative and qualitative results discussed in Chapter 6 and 7 suggest that SEA
has indeed succeeded in this objective.
8.6.1.4 Is SLA like L1 Acquisition?

In Section 3.3.3, in Chapter 3, references to Bley-Vroman’s argument for a set of domain-general cognitive procedures for L2 learning in the place of the domain-specific LAD operating in L1 acquisition (Bley-Vroman, 1989) were made. It can be hypothesised that to expand the rhythmic hierarchy of L1 to include L2 rhythms, procedures used in L1 might be equally useful for language learning. In the case of L2 Mandarin learning, we simply do not know enough about the audition process to even contemplate answering the question about whether learning Mandarin as L1 is the same as learning Mandarin as L2.

8.6.1.5 Future research

The present study has documented the progress of two groups of adult English speakers during their initial stage of learning Mandarin prosody using SEA. The major difficulties experienced by students have been identified and the primary cause for these has been identified as the interference of English intonation patterns. In general, these difficulties are in agreement with observations reported by other researchers in this field.

This study demonstrated that applying SEA to the teaching of Mandarin prosody appeared to be motivational and effective for the experimental group of students. In particular, though the two groups differed in their average age, the older group of students (experimental group) seemed to outperform the younger cohort in the control group. However, due to the small sample size of students involved, no conclusive evidence has been obtained to resolve that question of whether age negatively affects the learning of L2. In future studies, it might be worth investigating whether age helps in a learning environment such as SEA. What is the explanation for the superior performance of older students? Could it be because they worked harder in SEA? If so,
what caused them to work harder in SEA? A similar experiment involving larger sample sizes using SEA might provide some answers to these questions.

Another possible extension of this study would be a longitudinal investigation of how the Mandarin pronunciation ability of students taught with SEA improves over periods longer than one semester.

This study has brought out clearly the need to further consolidate student learning through the addition of a reinforcement session based on the patterns of errors obtained. A possible and desirable extension of this study would be to investigate how students taught with SEA can be helped with a range of computer enhanced learning programs such as Sptool (Zhang & Newman, 2003) other computer programs or activities targeting specific problem areas in acquiring Mandarin.

The relationship between the four macroskills of speaking, listening, reading and writing needs to be reexamined with regards to the teaching of Mandarin. The interrelatedness of the four macro skills suggests that the development of the skills should be integrated rather than separate. Though not part of this study, it is probable that students learning with SEA would have better listening comprehension skills as well as syntactic knowledge. This indeed can be another area for research.

The application of SEA to language learning is obviously not restricted to the teaching of Mandarin only. In fact, it is applicable to the teaching of any other language. Adult learners learning any language will find some areas of the L2 language’s phonology difficult. Therefore, with the appropriate enrichment of input and learning tools in the learning environment, this approach can be applied to the learning of any language. For instance, Asian L2 learners of
English tend to have trouble with consonant clusters at the beginning and end of English words. No doubt research findings and procedures in this study and other studies using the Verbotonal method can further contribute to the search for a solution for such problems.

Finally, it is encouraging to have found support for the use of exaggeration and feedback in teaching pronunciation in a number of recent research studies. McCandliss et al’s investigation on the success and failure in teaching the [r]-[l] contrast to Japanese adults is such an example. However, these studies investigated the effect of training on segmental contrasts such as [r]-[l]. The effect of training in overcoming the differences between English and Mandarin at the suprasegmental level using sentences requires further investigation. However, efforts in this area are hampered by not knowing what to focus the research on. While a lot has been done in comparative phonetics, little information is available in equal measure in comparative suprasegmental studies in cross-linguistic studies. Therefore, in my opinion, such similar research in the suprasegmental area between different languages should be carried out in future in tonal and non-tonal languages.

8.7 Concluding Remarks

The most important conclusion demonstrated by the findings of this research is that as designers of language programs, providing a list of good strategies that learners should adopt might be less effective than providing a learning environment which contains a large variety of “devices”, both linguistic and non-linguistic, to stimulate all the senses in our bodies to learn. It is arguable that the findings of this research demonstrate that perhaps the most effective learning machine is non-linguistic and is based in our own body. The power of SEA to teaching Mandarin prosody is a result of extensive and regular stimulation of the students’ multi-sensory perceptual systems.
This thesis set out to test a multi-sensory approach to teach Mandarin prosody and tones. This thesis can only partially describe the visible signs of learning. Many invisible aspects of learning were not and/or could not be open to description and may never be visible or even amenable to description as they are potentially hidden under many layers of complex individual behaviour patterns. It must be borne in mind that when the behaviour of a particular aspect was singled out for description (such as errors in Tone 3), in reality, that aspect is never isolated.

This thesis has found several important insights about the process of teaching Mandarin to English speaking adult learners. For instance, the issue of sequencing teaching materials is examined. The conclusion is that no particular sequencing of the teaching material is needed if learners are provided with easy access to a large amount of TL materials; and if learners are taught appropriate ways of dealing with such materials. Another finding is that learning the prosody of sentences of Mandarin is likely to be more effective for the acquisition of tones and a whole host of other prosodic phenomena in Mandarin. This is because by working on sentences and rhythm of the language, students physically but unconsciously realign their proprioception through the soma, skin, the sensation they feel when they move in the pit of their stomach and through the vibration they feel in their ears and bones. The results of this study show that experimental group of students also unconsciously adjusted their voice range to accommodate for the prosody of Mandarin when speaking Mandarin. This made their Mandarin speech more intelligible to NSs of Mandarin.

Through detailed analysis of tonal errors, it is also found that English students learning Mandarin tended to have poor control of their voice ranges especially with respect to Tone 1, Tone 2 and Tone 4. However, through focusing on sentences thus working on the realistic realizations of Tone 3, some learners have demonstrated partial mastery of Tone 3 in mid-
sentence positions even after only 30 hours of face-to-face contact. This suggests that although previous researchers’ suggested that Tone 3 should be taught either as low dipping (Lin, 1985) or half third (Hsieh, 1996), changing the context in which we present these tones might be just as effective.

Discussion on sequencing also raises questions about some of the processes (such as what constitutes comprehensible input; developmental sequence; order of difficulty in tone acquisition) that have influenced SLA and the field of teaching Mandarin as a L2 for decades. This thesis seems to suggest that in creating an effective language learning environment, what is important is not the search for the ultimate sequencing formula. Rather it is about creating a set of conditions which enable the learners to meet their personal learning needs. This thesis shows that when learners are allowed opportunities to follow their goals, they make significant progress in their learning even in a short time. Furthermore, by broadening the support systems to include the body, movement, gesture, and other forms of technical support, learners are supported in various ways. More importantly, through using these support systems, learners have learned new strategies for L2 learning thus transforming the way they learn.
Appendices

Appendix One: Sample teaching material

Week 1: INTRODUCTION

我来介绍介绍

After this week’s materials, you should be able to:
1. introduce yourself and others;
2. say hello and goodbye appropriately;
3. give people your name and surname in speech and in writing;
4. demonstrate the four tones in spoken Chinese correctly.

TASK 1
Listen to the five dialogues. Write the dialogue number to the correct picture.
(audiofile: w1 conversations)

Work out the meaning of the conversations in English by reading the characters.

Conversation 1:
A: 早, 我叫张明 (F)
B: 我是 Chris Bailey. 认识你我很高兴(F)
A: 认识你我也很高兴 (F)

Conversation 2:
A: 你好, 您贵姓? (F)
B: 我姓 Davies, 你呢? (F)
A: 我姓刘, 你叫什么名字? (F)
B: 我叫玛丽, 你呢? (F)
A: 我叫天星. 认识你我很高兴 (F)
B: 认识你我也很高兴. (F)

Conversation 3:
安娜: 伍爱华, 你好! (F)
伍爱华: 安娜, 您好. (F) 我给你们介绍介绍. (F) 这位是李先生. 这位是李太太. 这是李小姐. (F)
李先生, 李太太, 李小姐: 你好. (F)
安娜: 你们好! 认识你们也我很高兴. (F)
Conversation 4:
安娜: 李小姐, 你叫什么名字? (F)
李小姐: 我叫丽丽. (F)

Conversation 5:
安娜: 你好吗? (F)
李小姐: 我很好, 你呢? (F)
安娜: 我也很好, 谢谢. (F) 还好 (F) 马马虎虎. (F)

TASK 2
Listen and practise each conversation on the tape as closely to the models on the tape as possible.
Activity 1:
Get the students in a circle, scrunch up a piece of used paper into a ball and play throwing the ping pong ball. For example: the teacher selects one student and then says good morning and I am_____. It’s nice meeting you. Then, throws it to a student, the student says “Good morning, I am....... It’s nice meeting you all too.” Then throws it to another student and so on.

Activity 2:
Get all the names of the students in the class and then use the yes and no question form to confirm their identity. So,
A starts with: 请问, 你是..... 吗?
B: 是.
Then A introduces her/himself, then say it’s nice to meet you and then say goodbye. (Use the language in conversation 1 as a guide). Do this until all students’ and the teacher’s identity has been confirmed.

TASK 3:
Write the characters in a notebook until you can remember them.
Vocabulary 生词:
* 生: good morning
* 是: to be verb is, am are
* 早: come,
* 你: you [singular], pronoun
* 你们: you[plural], pronoun
* 您: you [singular] polite term for an older person
* 好: good [adjective]
* 我: I/me, pronoun
* 贵: expensive [adjective], honourable
* 姓: to be surnamed, verb
* 叫: to call [verb], to be called
* 什么: what
*名字: name [noun]
介绍: to introduce someone
先生: Mr.
太太: Mrs, a title (Mrs) or the colloquial form of wife
小姐: Miss
*这: this
位: polite measure word/counter for people
认识: to know of or to recognize someone by sight
*很: very adverb
高兴: happy [adjective]
*也: also [adverb]
*再: again [adverb]
*见: see [verb]
*他: he/him
*她: she/her
到: to come, to arrive
北京: Beijing.

The four tones:
Chinese is a language with different tones that are capable of differentiating meanings. A syllable, when pronounced in a different tone, has a different meaning even if it is composed of the same initial and final. In the Beijing dialect there are four basic tones represented respectively by the following tone-numbers:
1=第一声: 1st tone,
2=第二声: 2nd tone,
3=第三声: 3rd tone,
4=第四声: 4th tone.
Tone changes:
A 3rd tone, when immediately followed by another 3rd tone, should be pronounced in the 2nd tone.

Grammar: 语法

1. What is a yes and no question?
A yes and no question is a confirmation question by which one attempts to confirm one’s understanding by requesting an answer. The answer to a yes or no question can only be yes or no.
In English:
e.g. Are you Andrew? Yes, I am Andrew.
Note that in forming the question, the subject “you” and the verb to be (“are”) have been inverted in position in the question form. In English, the verb to be (“are”) precedes the subject (“you”) in questions.
Intonation pattern: Are you Andrew? Italicized syllable is stressed.
Start low, level, high on “are” then glide off.
In Mandarin, the yes or no question is simply formed by adding a question particle “ma”, to a statement.

* e.g. 我是 Andrew. ---＞ 你是 Andrew 马？

Intonation pattern: Each syllable maintains its tone without any change. In the question form, the ending “ma0” ends slightly higher.

是 is the verb to be for the first person, second person and the third person.

2. 是: the verb to be has a restricted use.
While in English, the verb to be (“is, am and are”) can be used in a variety of ways such as in e.g. I am going to China tomorrow.

是 (shi) in Mandarin is like an “equal” sign. The condition for using an equal sign in mathematics is that the things on the left-hand-side and the right-hand-side of the equal sign be completely identical in properties and in kind. Therefore, why do you think that in Mandarin, 我是 Andrew.” is permissible and

我是很好 is not, in which “很好” is an adverb or an adjective.

3. Negation
The negative marker “不” goes before a verb in Mandarin rather than after as in English. e.g. 我不是 Andrew.

The tone of “不=not” is usually a falling tone, except when the following word is in the falling tone, then “不” changes to “a rising tone.

* e.g. in 不是

4. Answer to a yes or no question:
There is no one word for “yes” or “no” in Mandarin. To answer “yes” to a question, one uses the affirmative form of the verb; and to answer “no”, one uses the negative form of the verb in the original question.

* e.g. 你是 Andrew 吗？

yes. 是。

no. 不是

5. Question particles: ma, ne.
In this week’s work, both question particles 吗 and 呢 have been introduced. While 吗 has been used to form a “yes/no question”, 呢 has been used as an equivalent to “...and you” in English and therefore is only used after a question has already been previously asked. For instance,

1. 我叫 David, 你呢？

2. 你是 Andrew 吗？

6. Adverbs
In this week’s work, we have introduced 也=also as the adverb. An adverb in Chinese comes immediately before the verb and can only be separated from it by another adverb. They cannot stand on their own without a verb.
**Week 1: 汉字：我来介绍介绍**

<table>
<thead>
<tr>
<th>汉字</th>
<th>意思: Try to guess the meaning first then confirm the answers with your teacher)</th>
<th>例子:</th>
</tr>
</thead>
<tbody>
<tr>
<td>你</td>
<td></td>
<td>你早！</td>
</tr>
<tr>
<td>早</td>
<td></td>
<td>早！</td>
</tr>
<tr>
<td>们</td>
<td></td>
<td>你们好！</td>
</tr>
<tr>
<td>您</td>
<td></td>
<td>您</td>
</tr>
<tr>
<td>好</td>
<td></td>
<td>你好！</td>
</tr>
<tr>
<td>我</td>
<td></td>
<td>我很好！</td>
</tr>
<tr>
<td>贵</td>
<td></td>
<td>这很贵.</td>
</tr>
<tr>
<td>再</td>
<td></td>
<td>再见</td>
</tr>
<tr>
<td>见</td>
<td></td>
<td>再见</td>
</tr>
<tr>
<td>他</td>
<td></td>
<td>他们好吗？</td>
</tr>
<tr>
<td>她</td>
<td></td>
<td>她好吗？</td>
</tr>
</tbody>
</table>

**Homework:**
All homework should be handed in by the lecture in the following week.
1. Complete the character sheets provided.
2. Watch the video clip “您贵姓？” and then transcribe the words in characters and then translate into English.
Appendix Two: Content page of the course

Week 1: Introduction 我 wǒ 给 gěi 你 nǐ 们 men 介 jiè 绍 shào 介 jiè 绍 shào

Week 2: Where are you from? 你 nǐ 是 shì 从 cóng 哪 nǎ 儿 er 来 lái 的 de?

Week 3: How is the family? 家 jiā 里 lǐ 人 rén 好 hǎo 吗 ma?

Week 4: What do you do? 你 nǐ 做 zuò 什 shén 么 me 工 gōng 作 zuò?

Week 5: What are your hobbies? 你 nǐ 有 yǒu 什 shén 么 me 爱 ài 好 hào?

Week 6: Visiting friends 看 kàn 朋 péng 友 you

Week 7: Do you have….? 有 yǒu 没 méi 有 yǒu….?

Week 8: Shopping 买 mǎi 东 dōng 西 xī

Week 9: I am sorry, I have a date tonight. 对 duì 不 bù 起 qǐ, 我 wǒ 今 jīn 天 tiān 晚 wǎn 上 shāng 有 yǒu 约 yuē.

Week 10: Making Telephone calls 打 dǎ 电 diàn 话 huà

Week 11: What do you do everyday? 你 nǐ 每 měi 天 tiān 做 zuò 什 shén 么 me?

Week 12: The weather 天 tiān 气 qì
Appendix Three: Teaching materials covered in the first six weeks

Week 1:
1. 早 zǎo, 我 wǒ 叫 jiào 张 zhāng 明 míng.
2. 我 wǒ 是 shì 基 Chris Bailey. 认 rěn 识 shí 你 nǐ 我 wǒ 很 hěn 高 gāo 兴 xìng.
3. 认 rěn 识 shí 你 nǐ 我 wǒ 也 yě 很 hěn 高 gāo 兴 xìng.
4. 你 nǐ 好 hǎo, 您 nín 贵 guì 姓 xìng?
5. 我 wǒ 姓 xìng Daviess, 你 nǐ 呢 ni?
6. 我 wǒ 姓 xìng 刘 liú, 你 nǐ 叫 jiào 什 shén 么 me 名 míng 字 zì?
7. 我 wǒ 叫 jiào 玛 mǎ 丽 lì, 你 nǐ 呢 ni?
8. 我 wǒ 叫 jiào 天 tiān 星 xīng, 认 rěn 识 shí 你 nǐ 我 wǒ 很 hěn 高 gāo 兴 xìng.
9. 认 rěn 识 shí 你 nǐ 我 wǒ 也 yě 很 hěn 高 gāo 兴 xìng.
10. 我 wǒ 爱 ài 华 huá, 你 nǐ 好 hǎo!
11. 您 nín 好 hǎo. 我 wǒ 给 gěi 你 nǐ 们 men 介 jiè 绍 shào 介 jiè 绍 shào. 这 zhè 位 wèi 是 shì 李 lǐ 先 xiān 生 shēng. 这 zhè 位 wèi 是 shì 李 lǐ 太 tài 太 tài. 这 zhè 是 shì 李 lǐ 小 xiǎo 姐 jiè.
12. 你 nǐ 们 men 好 hǎo! 认 rěn 识 shí 你 nǐ 们 men 也 yě 我 wǒ 很 hěn 高 gāo 兴 xìng.
13. 李 lǐ 小 xiǎo 姐 jiè, 你 nǐ 叫 jiào 什 shén 么 me 名 míng 字 zì?
14. 李 lǐ 小 xiǎo 姐 jiè: 我 wǒ 叫 jiào 丽 lì 丽 lì.

Week 2:
15. 你 nǐ 好 hǎo 吗 ma?
16. 我 wǒ 很 hěn 好 hǎo, 你 nǐ 呢 ni?
17. 我 wǒ 也 yě 很 hěn 好 hǎo, 谢 xiè 谢 xiè. / 还 huán 好 hǎo/ 马 mǎ 马 mǎ 虎 hǔ 虎 hǔ.
18. 他 tā/ 她 tā 叫 jiào ____________.
19. 他 tā/ 她 tā 是 shì 从 cóng 澳 ào 大 dà 利 lì 亚 yà 来 lái 的 de.
20. 你 nǐ 是 shì 从 cóng 哪 nǎ 儿 er 来 lái 的 de?
21. 我 wǒ 是 shì 从 cóng 澳 ào 大 dà 利 lì 亚 yà 来 lái 的 de.
22. 忙 máng 吗 ma?
23. 很 hěn 忙 máng.
24. 好 hǎo.
25. 我 wǒ 先 xiān 走 zǒu 了 le.
26. 再 zài 见 jiàn.
27. 张 zhāng 老 lǎo 师 shī, 吃 chī 饭 fàn 了 le 吗 ma?
28. 还 huán 没 méi 吃 chī 呢 ni, 你 nǐ 呢 ni(níne)?
29. 吃 chī 了 le,
30. 一 yī 会 huì 儿 er 见 jiàn.
31. 同 tóng 学 xué 们 men 好 hǎo.
32. 老 lǎo 师 shī 好 hǎo.
33. 老 lǎo 师 shī 早 zǎo,
34. 还 huán 可 kě 以 yǐ.
35. 有 yǒu 事 shí 儿 er 吗 ma?
36. 有 yǒu.
37. 有 yǒu 什 shén 么 me 事 shí 儿 er?
38. 请 qǐng 问 wèn, 这 zhè 个 gè 字 zì 怎 zěn 么 me 写 xiě?
39. How do you read “zēn”?
40. How do you pronounce “shēng”?
41. How do you pronounce “shēng”?
42. What do you mean by “shé”?
43. What does “xiè” mean, and how do you pronounce it?

Week 3:
44. Who is this? Is this person shì? Who is he?
45. Who is that? Is this person shì? Who is she?
46. Who is his wife? Is this person shì? Who is he?
47. Have you got a boy/girl friend? You nǐ有 you/女 nü3 朋 péng 友 yǒu/男 nán 朋 péng 友 yǒu 吗 ma?
48. Who is he? Is this person shì? Who is she?
49. Do you have any brothers and sisters? You nǐ有 yǒu 兄 xiōng 弟 dì 妹 mèi 吗 ma?
50. Is this your (older) brother? Yes./No. Is this your (older) brother? Yes./No. Is this person shì? Who is she?
52. Do you have any children? Yes/No. You nǐ有 yǒu 孩 hái 子 zǐ 吗 ma? 有 yǒu/没 méi 有 yǒu.
53. How many children do you have? You nǐ 有 yǒu 几 jǐ 个 gé 孩 hái 子 zǐ?
54. Are they female or male? Is this person shì? Are they (female/male)?
55. Is he married to her? He tā 是 shì 她 tā 工 gōng 作 zuò?
56. Is this your father? He tā 是 shì 他 tā 大 dà 家 jiā 人 rén?
57. Yes. 对 duì.
58. Incorrect. 不 bù 对 duì.
59. How old is your father? You nǐ 爸 bā 爸 bà 多 duō 大 dà 年 nián 纪 jì 了 le?
60. How old are you? [for people over ten years old] You nǐ 多 duō 大 dà 了 le?
61. How old are you? [for kids younger than 10 years old] You nǐ 几 jǐ 岁 suì 了 le?
62. How many people are there in your family? You nǐ 家 jiā 里 lǐ 有 yǒu 几 jǐ 口 kǒu 人 rén?

Week 4:
63. What does your job? You nǐ 干 gàn 什 shén 么 me 工 gōng 作 zuò?
64. My profession is........ 我 wǒ 是 shì........
65. Where do you work? I work at Hyatt hotel. You nǐ 在 zài 哪 nǎ 儿 er 工 gōng 作 zuò? 我 wǒ 在 zài Hyatt 酒 jiǔ 店 diàn 工 gōng 作 zuò.
66. Do you work at a restaurant? Yes, you nǐ 在 zài 饭 fàn 馆 guǎn 工 gōng 作 zuò 吗 ma? 对 duì.
68. Which university? University of Canberra. 哪 nǎ 个 ge 大 dà 学 xué? 堪 kān 培 péi 拉 lá 大 dà 学 xué.
70. What languages can you speak? English, French and a little Chinese, a little Japanese.
103. I would like to invite you to a meal.
102. I will make the tea.
100. (Don’t stand on ceremony.)
98. Thank you.
97. I am very thirsty, can I have a beer?
96. I want a bottle of beer./I do not drink beer.
92. Please wait a moment.
94. I am sorry, I am late.
90. Please sit down.
89. Week 6:
88. 你
87. 你
86. 你
85. 你
84. 你
83. 你
82. 你
81. 你
80. 你
79. Do you have a mobile phone?
78. Do you have email?
77. What’s your fax number?
76. What’s your phone number?
75. What’s your address?
74. Where do you live?
73. Please let me introduce myself.
72. Do you mind if I ask you a few questions? No.
71. Do you have a mobile phone? You have a mobile phone? Have you a mobile phone? Have you got a mobile phone? Have you got mobile phone? Have you got mobile phone?
70. Week 5:
69. 你
68. 在
67. 哪
66. 那
65. 儿
64. 人
63. 工
62. 作
61. 作
60. 作
59. 作
58. 作
57. 作
56. 作
55. 作
54. 作
53. 作
52. 作
51. 作
50. 作
49. 作
48. 作
47. 作
46. 作
45. 作
44. 作
43. 作
42. 作
41. 作
40. 作
39. 作
38. 作
37. 作
36. 作
35. 作
34. 作
33. 作
32. 作
31. 作
30. 作
29. 作
28. 作
27. 作
26. 作
25. 作
24. 作
23. 作
22. 作
21. 作
20. 作
19. 作
18. 作
17. 作
16. 作
15. 作
14. 作
13. 作
12. 作
11. 作
10. 作
9. 作
8. 作
7. 作
6. 作
5. 作
4. 作
3. 作
2. 作
1. 作
0. 作

Please sit down. Please come in. Please wait a moment. Please forgive me. I am sorry, I am late. I want a cup of tea/water/coffee. I want a bottle of beer. I do not drink beer. I am very thirsty, can I have a beer? I want to invite you to dinner. I will make the tea. I would like to invite you to a meal.

…公 gōng 司 sī 工 gōng 作 zuò.
82. 你 nǐ 喜 xǐ 欢 huan 什 shén 么 me?
83. 你 nǐ 有 yǒu 什 shén 么 me 爱 ài 好 hǎo?
84. 我 wǒ 喜 xǐ 欢 huan 看 kàn 书 shū, 听 tīng 音 yīn 乐 yuè, 聊 liáo 天 tiān 尔 er, 最 zuì 讨 tǎo 厌 yàn 看 kàn 电 diàn 视 shì.
85. 你 nǐ 喜 xǐ 欢 huan 什 shén 么 me 爱 ài 好 hǎo?
86. 我 wǒ 喜 xǐ 欢 huan 做 zuò 饭 fàn.
87. 我 wǒ 不 bù 太 tài 喜 xǐ 欢 huan 做 zuò 饭 fàn.
88. 我 wǒ 不 bù 太 tài 喜 xǐ 欢 huan 吃 chī 饭 fàn.
89. 我 wǒ 不 bù 太 tài 喜 xǐ 欢 huan 想 xiǎng 什 shén 么 me 运 yùn 动 dòng?
90. 请 qǐng 坐 zuò.
91. 请 qǐng 进 jìn.
92. 请 qǐng 等 děng 一 yī 等 děng.
93. 请 qǐng 原 yuán 谅 liàng 我 wǒ.
94. 我 wǒ 想 xiǎng 请 qǐng 你 nǐ 吃 chī 饭 fàn.
95. 我 wǒ 想 xiǎng 喝 hē 饮 yǐn 水 shuǐ/咖 kā 啡 fēi.
96. 我 wǒ 想 xiǎng 做 zuò 饭 fàn.
97. 我 wǒ 想 xiǎng 想 xiǎng 喝 hē 饮 yǐn 水 shuǐ/咖 kā 啡 fēi 吗 ma?
98. 谢 xiè 谢 xiè.
99. 你 nǐ 不 bù 用 yòng 谢 xiè.
100. (Don”t stand on ceremony.) 不 bù 客 kè 气 qì.
101. 欢 huān 迎 yíng, 欢 huān 迎 yíng.
102. 我 wǒ 将 jiāng 做 zuò 茶 chá.
Appendix Four: Guiding Questions for Interview

Student background

1. Do you work and study during the semester?

2. How many hours do you study per week?

The learning process

1. In terms of the study you do, what kind of activities do you do?

2. Do you do any preparation before you come to class? If yes, what kind of preparation?

3. What activities in class did you find useful during the class for learning?

4. Did you use any of the activities you did in class in your own learning?

5. Should we have more video materials next semester?

6. Did you use the speech analysis tool? If yes, how did you use it? If no, why not.

How do you relate to tones?

7. In terms of the CDs, what did you use?

8. Do you have a Mandarin tutor? How often do you see your tutor and what do you do in those sessions?

9. How do you feel about using Pinyin right from the beginning?

10. In terms of the classes in the second semester, what improvements would you like to suggest?

<table>
<thead>
<tr>
<th>Memory</th>
<th>Cognitive</th>
<th>Compensation</th>
<th>Metacognitive</th>
<th>Affective</th>
</tr>
</thead>
<tbody>
<tr>
<td>Creating mental images</td>
<td>Practicing</td>
<td>Guessing intelligently</td>
<td>Centering your learning</td>
<td>Lowering your anxiety</td>
</tr>
<tr>
<td>- grouping</td>
<td>-repeating</td>
<td>-using linguistic clues</td>
<td>-overviewing/</td>
<td>-using progressive relaxation</td>
</tr>
<tr>
<td>- associating/</td>
<td>-formally practising sounds</td>
<td>-using other clues</td>
<td>linking with already</td>
<td>and deep breathing</td>
</tr>
<tr>
<td>- elaborating</td>
<td>and writing systems</td>
<td>Overcoming limitations in</td>
<td>known material</td>
<td>-using music</td>
</tr>
<tr>
<td>- placing new words in</td>
<td>-recognising and using</td>
<td>speaking and writing</td>
<td>-paying attention</td>
<td>-using laughter</td>
</tr>
<tr>
<td>context</td>
<td>patterns</td>
<td>-switching to L1</td>
<td>-delaying speech to</td>
<td></td>
</tr>
<tr>
<td>Applying images and</td>
<td>-recombining</td>
<td>-getting help</td>
<td>concentrate on listening</td>
<td></td>
</tr>
<tr>
<td>sounds</td>
<td>-practising naturalistically</td>
<td>-using mime or gesture</td>
<td></td>
<td></td>
</tr>
<tr>
<td>- using imagery</td>
<td>-use multimedia resources</td>
<td>-avoiding communication</td>
<td></td>
<td></td>
</tr>
<tr>
<td>- semantic mapping</td>
<td>-use interactive</td>
<td>-selecting the topic</td>
<td></td>
<td></td>
</tr>
<tr>
<td>- using key words</td>
<td>conversations</td>
<td>-adjusting or</td>
<td></td>
<td></td>
</tr>
<tr>
<td>- representing sounds in</td>
<td>Receiving and sending</td>
<td>approximating the</td>
<td></td>
<td></td>
</tr>
<tr>
<td>memory</td>
<td>messages</td>
<td>message</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reviewing well</td>
<td>- getting the idea quickly</td>
<td>-coining words</td>
<td></td>
<td></td>
</tr>
<tr>
<td>- structural reviewing</td>
<td>-using resources for</td>
<td>-using circumlocution</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Employing action</td>
<td>receiving and sending</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>- using physical</td>
<td>messages</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>response</td>
<td>Analysing and reasoning</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>- using mechanical</td>
<td>-reasoning deductively</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>techniques</td>
<td>-analysing expression</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>-analysing contrastively</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>- translating</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>- transferring</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Creating structure for input</td>
<td>Creating structure for input</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>and output</td>
<td>and output</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>- taking notes</td>
<td>- taking notes</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>- summarising</td>
<td>- summarising</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>- highlighting</td>
<td>- highlighting</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Direct (performance)  Indirect (direction)
Appendix Six: Transcription of a conversation between Student 1 and a Thai student
from experimental group

Duration: 11m 7s

Fiat: เข้าถึงให้ คุณนั่งอยู่
Peter: ขอบคุณค่ะ
Fiat: คุณเข้าถึง หรือไม่?
Peter: คุณเข้าถึง หรือไม่
Fiat: คุณเข้าถึง หรือไม่
Peter: คุณเข้าถึง หรือไม่
Fiat: คุณเข้าถึง หรือไม่
Peter: คุณเข้าถึง หรือไม่

Fiat: คุณรู้จักมีเพื่อนใช้ไหม?
Peter: คุณรู้จักมีเพื่อนใช้ไหม?
Fiat: คุณรู้จักมีเพื่อนใช้ไหม?
Peter: คุณรู้จักมีเพื่อนใช้ไหม?
Fiat: คุณรู้จักมีเพื่อนใช้ไหม?
Peter: คุณรู้จักมีเพื่อนใช้ไหม?

Fiat: คุณรู้จักมีเพื่อนใช้ไหม?
Peter: คุณรู้จักมีเพื่อนใช้ไหม?
Fiat: คุณรู้จักมีเพื่อนใช้ไหม?
Peter: คุณรู้จักมีเพื่อนใช้ไหม?
Fiat: คุณรู้จักมีเพื่อนใช้ไหม?
Peter: คุณรู้จักมีเพื่อนใช้ไหม?

Fiat: คุณรู้จักมีเพื่อนใช้ไหม?
Peter: คุณรู้จักมีเพื่อนใช้ไหม?
Fiat: คุณรู้จักมีเพื่อนใช้ไหม?
Peter: คุณรู้จักมีเพื่อนใช้ไหม?
Fiat: คุณรู้จักมีเพื่อนใช้ไหม?
Peter: คุณรู้จักมีเพื่อนใช้ไหม?

Fiat: คุณรู้จักมีเพื่อนใช้ไหม?
Peter: คุณรู้จักมีเพื่อนใช้ไหม?
Fiat: คุณรู้จักมีเพื่อนใช้ไหม?
Peter: คุณรู้จักมีเพื่อนใช้ไหม?
Fiat: คุณรู้จักมีเพื่อนใช้ไหม?
Peter: คุณรู้จักมีเพื่อนใช้ไหม?
Peter: 他 tā 有 yǒu 一 yī 个 gè 弟 dì 弟 dì 叫 jiàoHUGH, 他 tā 是 shì27 岁 sui 还 huán 没 méi 有 yǒu 结 jié 婚 hun, 他 tā 是 shì 大 dài 学 xué 生 shēng, 他 tā 会 huì 说 shuō 英 yīng 语 yǔ, 他 tā 喜 xǐ 欢 huan 骑 qí 车 chē.
Fiat: 他 tā 会 huì 说 shuō 儿 ěr 种 zhǒng 语 yǔ 言 yán．
Peter: 啊啊 a, 他 tā 会 huì 说 shuō 英 yīng 语 yǔ, 还 hái 有 yǒu 一 yī 个 gè 弟 dì 弟 dì, 叫 jiàoPATRICK, 他 tā 是 shì21 岁 sui, 他 tā 也 yě 是 shì 大 dài 学 xué 生 shēng, 他 tā 会 huì 说 shuō 英 yīng 语 yǔ, 他 tā 住 zhù 在 zài SYDNEY.
Fiat: 我 wǒ 大 dà 妹 mèi 妹 mèi
Peter: 啊啊 a 我 wǒ 有 yǒu 一 yī 个 gè 妹 mèi 妹 mèi 叫 jiàoELISA, 他 tā 啊 a 他 tā 多 duō 大 dà 了 le? 他 tā32 岁 sui, 他 tā 结 jié 婚 hun 了 le, 他 tā 有 yǒu 三 sān 个 gè 孩 hái 子 zi, 两 liǎng 个 gè 儿 ěr 子 zi, 一 yī 个 gè 女 nǚ 儿 ěr.
Fiat: 好 hǎo 啊 a
Peter: 谢 xiè 谢 xiè, 他 tā 是 shì 一 yī 个 gè 学 xué… 小 xiǎo 学 xué 的 de 老 lǎo 师 shī, 他 tā 会 huì 说 shuō 英 yīng 语 yǔ, 他 tā 住 zhù 在 zài CAMBELL TOWN, 他 tā 喜 xǐ 欢 huan, 他 tā 喜 xǐ 欢 huan, 游 yóu 泳 yǒng.
Fiat: 大 dài 妹 mèi 妹 mèi 妹 mèi
Peter: 他 tā 也 yě, 他 tā 还 hái 有 yǒu 一 yī 个 gè 妹 mèi 妹 mèi 叫 jiàoMARY, 他 tā23 岁 sui, 没 méi 结 jié 婚 hun 了 le, 没 méi 结 jié 婚 hun, 没 méi 有 yǒu 结 jié 婚 hun. 他 tā 是 shì 一 yī 个 gè 中 zhōng 学 xué 的 de 老 lǎo 师 shī, 他 tā 会 huì 说 shuō 英 yīng 语 yǔ, 他 tā 住 zhù 在 zài CAMBELL TOWN.
Peter: 我 wǒ 很 hěn 喝 hē, 我 wǒ 涕 tì 一 yī 杯 bèi 水 shuǐ.
Fiat: 等 děng 一 yī 等 děng 啊 a,
Peter: 好 hǎo, 谢 xiè 谢 xiè, 谢 xiè 谢 xiè.
Fiat: 没 méi 关 guān 系 xi．
Peter: 忙 máng 啊 ma?
Fiat: 我 wǒ 很 hěn 忙 máng 啊 a, 好 hǎo 了 le, 我 wǒ 先 xiān 走 zǒu 了 le.
Peter: 好 hǎo, 你 nǐ 识 shí 你 nǐ 我 wǒ 很 hěn 高 gāo 兴 xīng,
Fiat: 你 nǐ 识 shí 你 nǐ 我 wǒ 也 yě 很 hěn 高 gāo 兴 xīng, 再 zài 见 jiàn
Peter: 再 zài 见 jiàn.
Conversation 2:
Peter: 你 nǐ 好 hǎo, 你 nǐ 好 hǎo
Fiat: 你 nǐ 好 hǎo
Peter: 你 nǐ 叫 jiào 什 shén 么 me 名 míng 字 zi
Fiat: 我 wǒ 叫 jiàoFIAT
Peter: 谢 xiè 谢 xiè, 我 wǒ 叫 jiàoPETER, 你 nǐ 不 bù 介 jiè 意 yì 我 wǒ 问 wèn 你 nǐ 一 yī 些 xiē 问 wèn 题 tí 吗 ma (你 nǐ 不 bù 介 jiè 意 yì 我 wǒ 问 wèn 题 tí 这 zhè 个 gè 问 wèn 题 tí 吧 ba)
Fiat: 问 wèn 吧 ba.
Peter: THAT’S OK, 谢 xiè 谢 xiè, THAT’S OK
Fiat: 问 wèn 吧 ba.
Peter: 好 hǎo 好 hǎo, 你 nǐ 多 duō 大 dà 了 le?
Fiat: 我 wǒ 多 duō 大 dà 了 le, 我 wǒ 多 duō, 22 岁 sui (sui).
Peter: 结 jié 婚 hun 了 le 吗 ma
Fiat: 没 méi 有 yǒu 啊 a.
Peter: 好 hǎo, 你 nǐ 干 gàn 什 shén 么 me 工 gōng 作 zuò.
Fiat: 我 wǒ 干 gàn WANNAIassa, 哦 è, 干 gàn 学 xué 生 shèng, 学 xué 生 shèng.
Peter: 好 hǎo 好 hǎo, 你 nǐ 会 huì 说 shuō 几 jǐ 种 zhǒng 语 yǔ 言 yán.
Fiat: 啊 a, 英 yīng 文 wén, 泰 tài 语 yǔ, 一 yī 点 diǎn 中 zhōng 语 yǔ YUN
Peter: 好 hǎo 好 hǎo 很 hěn 好 hǎo, 你 nǐ 往 zhù 哪 nà 儿 er?
Fiat: 我 wǒ 往 zhù WANNAIASSA.
Peter: 你 nǐ 是 shì 从 cóng 哪 nà 儿 er 来 lái 的 de.
Fiat: 我 wǒ 是 shì 泰 tài 国 guó 人 rén.
Peter: 你 nǐ 喜 xǐ 欢 huān 爱 ài 好 hǎo 吗 ma?
Fiat: 我 wǒ 喜 xǐ 欢 huān 游 yóu 泳 yōng, 做 zuò 饭 fàn 好 hǎo.
Peter: 好 hǎo 好 hǎo, 你 nǐ 爸 bà 爸 ba?
Fiat: 我 wǒ 爸 bà 妈 má 叫 jiàowULAN, 他 tā44 岁 suì, 他 tā 商 shāng 人 rén, 他 tā 喜 xǐ 欢
huān 唱 chàng 歌 gē, 喜 xǐ 欢 huān 卡 kǎ 拉 lā OK.
Peter: 啊 a 唱 chàng 歌 gē, 唱 chàng 歌 gē A SINGER, 好 hǎo.
Fiat: 好 hǎo 好 hǎo, 我 wǒ 妈 má 妈 má 叫 jiàoDU TAHUAN, 我 wǒ 妈 má 妈 má 家 jiā 庭 tíng 妇 fù 女 nǚ3, 她 tā45 岁 suì, 他 tā 喜 xǐ 欢 huān 做 zuò 饭 fàn, 我 wǒ
妹 mèi 妹 mèi…
Peter: 我 wǒ, 你 nǐ 爸 bà 爸 ba 妈 má 妈 má 是 shì 从 cóng 哪 nà 儿 er 来 lái 的 de?
Fiat: 我 wǒ, 他 tā 们 men 从 cóng 泰 tài 国 guó, 南 nán 部 bù.
Peter: 好 hǎo 好 hǎo, 谢 xiè 谢 xiè, 你 nǐ 妹 mèi 妹 mèi.
Fiat: 我 wǒ 妹 mèi 妹 mèi 叫 jiàoFRAN, 他 tā21 岁 suì, 他 tā 学 xué 生 shèng, 他 tā 喜 xǐ 欢
huān 吃 chī 饭 fàn.
Peter: 结 jié 婚 hun 了 le 吗 ma
Fiat: 没 méi 有 yǒu 啊 a, 大 dà 妹 mèi 妹 mèi 啊 a, NO, 你 nǐ 妹 mèi 妹 mèi 啊 a.
Peter: 你 nǐ 妹 mèi 妹 mèi, 你 nǐ 妹 mèi 妹 mèi, 你 nǐ 妹 mèi 妹 mèi 会 huì 说 shuō 几 jǐ 种
zhǒng 语 yǔ 言 yán.
Fiat: 他 tā 会 huì 说 shuō 泰 tài 语 yǔ, 英 yīng 语 yǔ.
Peter: 你 nǐ 弟 dì 弟 dì, 你 nǐ 一 yī 个 gè 弟 dì 弟 dì.
Fiat: 我 wǒ 弟 dì 弟 dì 叫 jiàoFIRST, 他 tā19 岁 suì, 他 tā 学 xué 生 shèng, 他 tā 喜 xǐ 欢 huān
打 dǎ 篮 lán 球 qiú, 他 tā 说 shuō, 会 huì 说 shuō 泰 tài 语 yǔ, 英 yīng 语 yǔ AND 两 liǎng 个
gè 弟 dì 弟 dì.
Peter: 还 hái 有 yǒu 一 yī 个 gè 弟 dì 弟 dì.
Fiat: YE 我 wǒ 弟 dì 弟 dì 叫 jiàoFRANK, 他 tā18 岁 suì, 他 tā 学 xué 生 shèng, 他 tā 喜 xǐ 欢
huān 游 yóu 泳 yōng, 他 tā 会 huì 说 shuō 泰 tài 国 guó, 泰 tài 语 yǔ, 英 yīng 语 yǔ, 其 qí 中
zhǒng 语 yǔ, 一 yī 点 diǎn 日 rì 本 běn.
Peter: 好 hǎo, 好 hǎo 很 hěn 好 hǎo, 他 tā 住 zhù 哪 nà 儿 er?
Fiat: 他 tā 住 zhù WANNAIASSA.
Peter: 好 hǎo 好 hǎo 谢 xiè 谢 xiè, 谢 xiè 谢 xiè
Fiat: 我 wǒ 很 hěn 喝 hè, 你 nǐ 我 wǒ 一 yī 杯 bēi 水 shuǐ 啊 a
Peter: 好 hǎo, 等 děng 一 yī 等 děng!
Appendix Seven: Transcription of a conversation between 2 students in the control group

R: is an Australian and TQ is a Vietnamese student whose data was not included in the analysis.

TQ: : 早 zǎo , 我 nǐ 好 hǎo 。
R: : 早 zǎo , 你 nǐ 好 hǎo , 你 nín 贵 guì 姓 xìng ？
TQ: : 我 wǒ 姓 xìng N 。你 nǐ 呢 ne ？
R: : 我 wǒ 姓 xìng B 。你 nǐ 叫 jiào 什 shén 么 me 名 míng 字 zi ？
TQ: : 我 wǒ 叫 jiào TQ ，你 nǐ 呢 ne ？
R: : 我 wǒ 叫 jiào R 。认 rèn 识 shí 我 wǒ 很 hěn 高 gāo 兴 xìng 。
TQ: : 认 rèn 识 shí 你 nǐ , 我 wǒ 也 yě 很 hěn 高 gāo 兴 xìng 。这 zhè 位 wèi 是 shì 谁 shéi ？
R: : 他 tā 是 shì 我 wǒ 爸 ba 爸 ba 。
TQ: : 他 tā 叫 jiào 什 shén 么 me 名 míng 字 zi ？
R: : 他 tā 叫 jiào J 。
TQ: : 那 nà 他 tā 太 tài 太 tai 是 shì 谁 shuí ？
R: : 他 tā 的 de 太 tài 太 tai 是 shì 她 tā 。
TQ: : 她 tā 叫 jiào 什 shén 么 me 名 míng 字 zi ？
R: : 她 tā 叫 jiào B 。
TQ: : 他 tā 们 men5 多 duō 大 dà 了 le ？
R: : 他 tā 们 men5 46 和 hé sìshí sān suì 了 le 。
TQ: : 他 tā 们 men5 住 zhù 哪 nǎ er ？
R: : 他 tā 住 zhù Q 和 hé 他 tā 住 zhù Q 。
TQ: : 他 tā 们 men5 从 cóng 哪 nǎ er 来 lái 的 de ？
R: : 他 tā 和 hé 他 tā P 。这 zhè 位 wèi 是 shì 谁 shuí ？
TQ: : 他 tā 是 shì 我 wǒ 爸 ba 爸 ba 。
R: : 他 tā 叫 jiào shénme 名 míng 字 zi ？
TQ: : 他 tā 叫 jiào B 。
R: : 那 nà 他 tā 的 de 太 tài 太 tai 是 shì 谁 shuí ？
TQ: : 他 tā 的 de 是 shì 她 tā 是 shì 她 tā 。
R: : 她 tā 叫 jiào shénme 名 míng 字 zi ？
TQ: : 她 tā 叫 jiào B 。
R: : 他 tā 们 men5 多 duō 大 dà 了 le ？
TQ: : 他 tā sìshí sān suì 和 hé sìshí sān suì 了 le 。
R: : 他 tā 们 men5 住 zhù 哪 nǎ er ？
TQ: : 他 tā 们 men5 住 zhù D 。
R: : 他 tā 们 men5 从 cóng 哪 nǎ er 来 lái 的 de ？
R: 对 duì.
TQ: 他 tā 叫 jiào shénme 名 míng 字 zi?
R: 他 tā 叫 jiào R.
TQ: 他 tā 多 duō 大 dà 了 le?
R: 他 tā 十 shí 六 liù.
TQ: 他 tā 爱 ài 好 hào 什 shén么 me?
R: 他 tā 喜 xǐ 欢 huan5 吃 chī 饭 fàn.
TQ: 他 tā 多 duō 大 dà 了 le?
R: 他 tā shíliù.
TQ: 他 tā 有 yǒu 没 méi 有 yǒu 女 nǚ 朋 péng you?
R: 不 bù 知 zhī 道 dào 这 zhè 位 wèi 是 shì 谁 shuí?
TQ: 他 tā 是 shì 我 wǒ 弟 dì 弟 dì.
R: 他 tā 叫 jiào 什 shén么 me 名 míng 字 zi?
TQ: 他 tā 叫 jiào T.
R: 他 tā 多 duō 大 dà 了 le?
TQ: 他 tā shíqì 岁 suì 了 le.
R: 他 tā 喜 xǐ 欢 huan5 什 shén么 me?
TQ: 他 tā 喜 xǐ 欢 huan5 唱 chàng 歌 gē. 他 tā 是 shì 你 nǐ 哥 gē 哥 gē 吗 ma?
R: 他 tā 是 shì 弟 dì 弟 dì.
TQ: 他 tā 叫 jiăoshénme 名 míng 字 zi?
R: 他 tā 叫 jiăo ST.
TQ: 他 tā 有 yǒu 没 méi 有 yǒu 男 nán 朋 péng you?
R: 没 méi 有 yǒu.
TQ: 他 tā 多 duō 大 dà 了 le?
R: 他 tā 二 èr 十 shí 二 èr 岁 suì 了 le.
TQ: 他 tā 爱 ài 好 hào 什 shén么 me?
R: 他 tā 喜 xǐ 欢 huan5 跑 pāo 步 bù.
TQ: 他 tā 在 zài 哪 nà 儿 ěr 工 gōng 作 zuò?
R: 他 tā 是 shì C 工 gōng 作 zuò. 这 zhè 位 wèi 是 shì 谁 shuí?
TQ: 他 tā 是 shì 弟 dì 弟 dì.
R: 他 tā 多 duō 大 dà 了 le?
TQ: 他 tā shíèr 岁 suì 了 le.
R: 他 tā 叫 jiăo 什 shén么 memíng 字 zi?
TQ: 他 tā 叫 jiăo T.
R: 他 tā 喜 xǐ 欢 huan5 什 shén么 me?
TQ: 他 tā 喜 xǐ 欢 huan5 吃 chī 饭 fàn. 这 zhè 位 wèi 是 shì 谁 shuí?
R: 他 tā 是 shì 妹 mèi 妹 mei.
TQ: 她 tā 叫 jiăo 什 shén么 me 名 míng 字 zi?
R: 她 tā 叫 jiăo J.
TQ: 他 tā 有 yǒu 没 méi 有 yǒu 男 nán 朋 péng you?
R: 不 bù 知 zhī 道 dào.
TQ: 他 多 大 了 le？
R: 他 shī bā.
TQ: 他 爱 好 什 么 me？
R: 他 喜 欢 书 shū．
TQ: 我 先 走 了 le．
R: 再 见 jiàn．
Appendix Eight: Chinese 1A: Language and Culture: End of semester 1, 2003 Evaluation questionnaire

This evaluation questionnaire has been designed to elicit, as much as possible, students' response about the overall organization of the subject and the teaching materials used in the subject. The questionnaire is anonymous and the data gathered in this survey are strictly confidential and will not affect the grades of any students. Thank you for your cooperation and your genuine input represents an invaluable contribution to the design of the subject.

Written by: Ms. Felicia Zhang, Chinese 1a subject convenor, 2003.

1. Sex: Male/ Female (please circle)
2. Nationality:
3. Have you studied Mandarin before you started this subject? Yes, No.
   If yes, for how long?
4. Do you speak any other Chinese dialects? If yes, which language(s)?
   Please fill in the following survey, giving as much information as possible:

Have you used any of the following computer materials in your study this semester? Please circle the ones you have and write down approximately how often or how much you have used them.

<table>
<thead>
<tr>
<th>Names of computer software and video materials</th>
<th>Yes/NO,(If no, give reasons for not using the materials)</th>
<th>How much? (hours/days per week)Approximately</th>
<th>Reasons for Not using the materials</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cl1a data cd</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cl1a audio cd 1 (w1to 8)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cl1a audio cd 2(w10-12 + with Tell me more CD)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mmbrowse Cd and songs</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Other video materials (specify)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chinese VCDs</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chinese movies</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
2. Please comment on the overall teaching and the teaching materials you have used in as much
detailed as possible especially the computer materials.

3. Please specify in what ways any of the materials have helped or not helped you in your study.
What suggestions would you make for this subject next year or for any Chinese courses in
general?

4. Which teaching materials listed in the above table should be included in the next semester's
teaching materials? Do you think the balance between speaking, listening, reading and writing
is adequate?

5. What other suggestions do you have for next semester?
References


Hsieh, D. H.-Y. (1996). *Acquisition of Mandarin Tones by Adult English Speakers: An Experimental Study of Problems Encountered by a Group of Australian University Students in Their First Semester of Language Training*. Unpublished PhD., University of Queensland, Publisher [St. Lucia, Qld.].


