Thus of every country and region, even the full advantage of the advancing however, we need to share data and better mutual communication and understanding of COCOSDA and Oriental COCOSDA. The history of collaboration between language processing in this part of the world for our future collaboration during publication, I would like to thank tremendous effort and devotion. I

Hiroya Fujisaki

Table of Contents

ONE: Introduction

TWO: Outline of some Asian languages

(1) Phonology, Phonetics, Prosody, and Orthography of Chinese
C-Y Tseng, A. J. Li
(2) Phonology, phonetics, and orthography of Hindi
S. Agrawal, K. K. Arora, S. Arora, K. K. Goswami
(3) Phonology, phonetics, prosody, and orthography of Indonesian
H. Rizo, A. A. Arman
(4) Phones, phonemes, prosodies, and writing system in Japanese
M. Dantsuji
(5) The phonology, phonetics, prosody, and orthography of Korean
S-J Moon
(6) The phonetics of Malay: from spelling to waveform
Zuraidah M. D., G. Knowles
(7) A Brief introduction to the Mongolian
I. Dava, E. Dashitseren, Y. Aral, O. Sambuuadorj
(8) Nepali: Phonetics, phonology and orthography
B. N. Regmi
(9) Thai phonology, phonetics, prosody, and orthography
S. Luksaneeyanawin
(10) The phonology, phonetics, prosody and orthography of Vietnamese
C-M Luong

THREE: Data centers of speech corpora

(1) The Chinese Linguistic Data Consortium (CLDC)
B. Xu, J. H. Tao
(2) The Chinese Corpus Consortium (CCC)
T. F. Zheng
(3) Language Resource Association (GSK)
S. Itahashi, K. Hasida, K. Matsui, H. Tanaka
(4) NII-Speech Resources Consortium (NII-SRC)
S. Itahashi, T. Ohsuga, K. Yamakawa
(5) ATR-NICT speech and language corpora
S. Nakamura
(6) Speech Information Technology and Industry Promotion Center (SITEC)
Y-J Lee
(7) The Association for Computational Linguistics and Chinese Language Processing (ACLCLP)
H-C Wang

FOUR: Speech corpora of Asian languages

4.1 Monologue
(1) CU2C : A dual-condition Cantonese speech database for speaker recognition
Y. Zheng, C. Qin, T. Lee, P-C Ching
(2) Text and speech corpora development in Indian languages
S. S. Agrawal, K. K. Arora, S. Arora, Samudravijaya K
(3) Speech corpora in Bahasa Indonesia
H. Rizo
1 INTRODUCTION
This paper outlines the phonetic patterns of the kind of standard Malay used by educated Malaysia in Kuala Lumpur, and discusses some of the theoretical problems involved in making a phonetic description. The set of phonemes presented here is extracted from the pronunciation entries in a working computer lexicon, and the phones referred to have emerged from the study of waveforms. The rules presented here correspond to rules in working computer programs.

2 THE PHONEMES OF MALAY
2.1 Vowels
The vowel system is a six term system, with five peripheral vowels and one central vowel (see also 2):

| Vowel | E | ə | U | a |

The writing system is generally phonemic, but fails to distinguish /e/ from /ə/, both are written with the letter “e”. Before a liquid or nasal, the vowel is most commonly /ə/, but there are a number of cases of homographic minimal pairs, e.g. semak /s?mak/ ‘overgrown’ and semak /semak/ ‘to check’. The close vowels /i, u/ have lowered and more central variants [I, U] which are used in the simplex form before a final consonant, e.g. nipis [nipIs] ‘thin’, cukup [cukUp] ‘enough’.

2.2 Consonants
In order to describe the consonant system we have to distinguish the basic set of native consonants from those introduced from other languages. The basic set includes the following:

<table>
<thead>
<tr>
<th>p, b</th>
<th>T, d</th>
<th>S</th>
<th>c, j</th>
<th>k, g</th>
<th>h</th>
</tr>
</thead>
<tbody>
<tr>
<td>m</td>
<td>n</td>
<td>l</td>
<td>ŋ</td>
<td>ŋ</td>
<td></td>
</tr>
<tr>
<td>(w)</td>
<td>y</td>
<td></td>
<td>w</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>