

Building a Unit Selection Speech Synthesiser for Malay Language Using FESTVOX and Hidden Markov Model Toolkit (HTK)

Wai Keong Loo^{1*}, Siew Hock Ow² and Roziati Zainuddin³

^{1,2}*Department of Software Engineering, Faculty of Computer Science & Information Technology, University of Malaya, Kuala Lumpur, Malaysia*

³*Department of Artificial Intelligence, Faculty of Computer Science & Information Technology, University of Malaya, Kuala Lumpur, Malaysia*

*Corresponding author. E-mail: david_loowaikeong@yahoo.com

ABSTRACT

This paper describes the development of unit selection speech synthesiser for Malay language by using the Festival speech synthesiser system. The development processes involved the study of the phonological nature of language, letter-to-sound modeling and clustering on the phone unit database. An HMM-based speech recognition tool kit and Hidden Markov Model Toolkit (HTK) were also applied to label the unit boundaries automatically. To evaluate the naturalness of speech, an overall quality test on the synthesisers was also conducted to compare the naturalness of the earlier Multi Band Resynthesis Overlap and Add (MBROLA) diphone concatenation synthesiser and the new FESTVOX unit selection synthesiser.

Key words: Unit Selection Speech Synthesiser; Malay Language; Letter-to-Sound Modeling; Hidden Markov Model Toolkit, HTK; MBROLA Diphone Concatenation; FESTVOX Unit Selection Synthesiser

INTRODUCTION

Until the year 2001, there was only one text-to-speech system in the Malay language, namely, SUM (*acronym for Sintesis Ucapan Melayu*) which was developed by researchers from the National University of Malaysia. The synthesiser model is based on the Klatt's formant synthesiser. However, there are quite a number of shortcomings on this synthesiser. The synthesised speech lacks naturalness, but yet no effort has been made on the high level analysis and improvement in prosody (Hussain et al., 1999). Owing to the demand of local telecommunication industry, there was indeed a need for a quality speech synthesiser in the Malay language. This initiated research into the Multi Band Resynthesis Overlap and Add (MBROLA) diphone concatenation approach (Dutoit et al., 1996) and unit selection approach (Hunt and Black, 1996).