PROSODIC ANALYSIS AND MODELLING FOR MALAY EMOTIONAL SPEECH SYNTHESIS

Mumtaz B. Mustafa[1,4], Raja N. Ainon[1,4*], Roziati Zainuddin[1,4*], Zuraidah M. Don[2,4*], Gerry Knowles[3,4] and Salimah Mokhtar[1,4*]

[1]Faculty of Computer Science and Information Technology, University of Malaya, Malaysia
[2]Faculty of Language and Linguistics, University of Malaya, Malaysia
[3]Lingenium Sdn Bhd, Kuala Lumpur, Malaysia
[4]Computational Speech Group, ICT Research Cluster, University of Malaya
mumshaka4@perdana.um.edu.my, *{ainon, roziati, zuraida, salimah}@um.edu.my,
g.knowles@lancaster.ac.uk

ABSTRACT

This paper discusses an emotional prosody generator for a Malay speech synthesis system that can re-synthesize the selected vocal emotion from neutral synthesized speech output and improve the naturalness by adopting rule-based prosody conversion techniques. The role of prosodic features in emotional expression, particularly fundamental frequency and duration, has been widely investigated in several research projects. This project attempts to improve the naturalness of the synthesized emotional Malay speech by establishing an effective mechanism for the re-synthesis of emotion. Such a mechanism is created by analyzing the variation in the F0 contour of continuous emotional Malay speech against a fixed time period. The emotional prosodic generator for Malay developed in the course of this research makes use of principles of parametric prosody manipulation to synthesize four basic emotions, namely happiness, anger, sadness and fear. Subjective evaluation by means of listening tests was conducted to validate the ability of the emotions generator to generate the necessary prosody to synthesize emotional expression. The evaluation results show an overall recognition rate of between 61% and 85%.

Keywords: Emotional speech re-synthesis; Prosody conversion; Rule-based approach, MBROLA

1.0 INTRODUCTION

Advances in speech synthesis research in recent years have made it possible to produce computer speech, which is very close to human speech. However, the generation of synthesized speech expressing emotions has been less successful, as it involves not only computer science but also other fields of speech research, including linguistics, psychology and signal processing.

One method of simulating emotions in neutral speech is by prosody modelling, in which a set of factors including F0, duration and intensity are controlled in order to convey non-lexical and pragmatic information in speech. Several studies have investigated the role of prosodic features in synthesized emotional speech in a range of languages and cultures, including [1], [2], [3] and [4]. This research focuses on improving the prosody modelling to address the issue of naturalness in synthesized emotional speech for Malay. Apart from some seminal studies [5], [6] and [7]; research on prosody and emotion for Malay TTS is still in its infancy. The objective of this research is to improve the naturalness of synthesized Malay emotional speech by investigating the prosodic properties of Malay speech and establishing an effective mechanism to synthesize a variety of emotional speech prosody using prosody manipulation techniques.

2.0 RELATED WORK

Various approaches have been proposed in the literature for the generation of appropriate prosody in synthesizing emotional speech, including corpus-based, rule-based, template-based and learning-based approaches. The rule-based approach simulates the emotion by manipulating speech properties such as F0, duration, intensity and voice quality. A rule-based approach of prosody conversion was developed for MBROLA by controlling the F0, duration and voice quality in [4].

A more recent learning approach such as HMM-based speech synthesis makes use of powerful statistical parameters to generate high quality synthesis with a much smaller voice database [8] and [9]. HMM-based speech synthesis also has the ability to generate emotional expression using its powerful statistical tool [10]. The problem is that HMM-based speech synthesis only generates arbitrary emotional expression [11]. Although there are a large number of prosodic features which can be used to analyse and describe variation in the prosody of speech, there are three features that are most consistently used for this purpose either singly or jointly [12], which are F0, duration and intensity.

Statistical analysis of F0, duration and intensity shows a strong correlation between the prosody and the emotion being expressed. Statistical analysis has been used by a number of researchers including Montero et al. [13] who investigate mainly duration and F0. Montero et al. [13] as well Murray and Arnott [2], simulate emotion using appropriate conversion factors for F0, duration and intensity. The former uses statistical analysis to derive the appropriate conversion rules while the latter uses direct observations of prosody in the speech contour. In this project, we investigate the prosodic properties corresponding to the four basic emotions happiness, anger, sadness and fear using the method proposed by Montero et al., for analyzing F0, duration and intensity at phoneme level. This method provides a clear indicator of the way emotional speech correlates with speech properties, and the analysis leads to the establishment of accurate conversion rules for prosody.

3.0 DATA COLLECTION AND THE ANALYSIS OF MALAY EMOTIONAL SPEECH

3.1 Malay Speech

Malay is non-tonal language spoken by more than 300 million people in several countries including Malaysia, Indonesia, Brunei, Singapore and Southern Thailand [14]. There are many dialects of Malay in Malaysia, but the focus in this study is on Standard Malay (SM), which is the variety of the language used in formal contexts such as education and the mass media [15]. The writing system of SM uses the standard Roman alphabet with 21 consonant letters and 5 vowel letters. Although this is the same alphabet as for English, the sound system of Malay differs from that of English [14]. The syllable pattern of any Malay word can have a combination of CVC (consonant vowel consonant), CV (consonant vowel), VC (vowel consonant) or just V (vowel) [16]. The most common syllable structures in SM are bi-syllabic and tri-syllabic, which together make up 97.52% of Malay words [16] and [17].

3.2 Recording Emotion in Speech

When designing the text material to be recorded, it is essential to ensure that sufficient prosodic information will be generated. It should be phonetically and prosodically balanced, and include a good coverage of the language [18]. The use of semantically neutral text makes it easier to compare emotional effects with each other while ensuring that the database is phonetically balanced. On the other hand, it also makes it more difficult for the speaker to express the emotion intended. By contrast, text that is semantically related to the intended emotion makes it easier for the speaker to express the emotion naturally.

For the purpose of developing the emotions generator, phonetically balanced sentences of different lengths and syllable structures were created. In all, 224 sentences of different lengths and syllable structures were used, 56 for each emotion (4 two-word sentences, 4 three-word sentences, 8 four-to nine-word sentences each). Each sentence is recorded by a female actor to express the different emotions, namely happiness, anger, sadness and fear. A single speaker is used in order to validate the effectiveness of the prosody generator [19] and to avoid problems of prosodic variation arising from the use of several voices. Table 1 lists some of the emotionally neutral sentences designed for the recording.

3.3 Evaluating the Expression of Recorded Malay Emotional Speech

Subjective evaluation is a valuable tool for the validation of the recorded speech as it demonstrates the ability of voice contributor to express emotion with high naturalness [18]. The emotional content of each recorded utterance was evaluated and validated by means of a listening test. There were twenty assessors, all native Malay speakers of different ages and from different professions. During the evaluation, each listener was required to listen to the voice snippet and then describe its emotional content. They were specifically asked to identify the emotion according to the prosody of the recording and not the semantic content of the sentence. The open choice method used for the evaluation gave a choice not only of the four emotions happiness, anger, sadness, fear, but also included an additional category 'others'. The option 'others' was added to allow the listeners to specify emotions that was not available in the list of choice provided.

Table 1. List of some sentences used for recording

Malay sentence	English translation
Sila duduk	*Please be seated*
Kamil belum bersedia	*Kamil is not ready*
Selangor akan menentang Kelantan	*Selangor will face Kelantan*
Sikap baik dipupuk dari kecil	*Good behavior is taught from small*
Kapal terbang milik kerajaan perlu diganti	*The government airplane must be replaced*
Benteng pertahanan Melaka telah dirobohkan oleh Portugis	*The fort of Melaka was destroyed by Portugal*
Remaja zaman ini terlalu manja dan tidak bertangggungjawab	*Teenagers of today are spoilt and have no sense of responsibility*

The listening test showed that the majority of native listeners can accurately identify the emotive elements in the recorded voice snippets, despite the lack of semantic meaning in some cases. Overall, anger has the highest rate of recognition with an average of 87.90%. Happiness was accurately identified at the rate of 84.40%. Sadness and fear had much lower rates at 70.50% and 77.80% respectively. Nevertheless, all four emotions have a recognition rate well above 70% on average. Table 2 presents the confusion matrix based on the results of the listening test.

Table 2. The confusion matrix of the listening test for the recorded speech

Recorded	Perceived as (%)								
Speech	Happy	Angry	Fear	Sad	Neutral	Excited	Worry	Bored	Disgust
Happy	**84.4**	5.7	2.5	2.2	1.8	3.4	0.0	0.0	0.0
Angry	4.3	**87.9**	2.1	2.9	0.3	0.0	0.0	0.0	2.5
Fear	3.1	2.2	**77.8**	7.2	3.3	0.0	6.4	0.0	0.0
Sad	3.9	2.4	14.3	**70.5**	2.3	0.0	4.2	2.4	0.0

3.4 Analysis of the Recorded Malay Emotional Speech

The analysis of the recorded utterances gives a general view of how the prosodic features differ from one emotion to another. The analysis was carried out using PRAAT [20]. The initial stage of analysis involved segmenting all 224 recorded utterances manually at phoneme level, and also marking silences and breathiness. Figure 1 illustrates the segmentation of an angry word using PRAAT.

The next stage of the analysis is the extraction of F0 for each phoneme. The F0 value at the beginning, the peak, the fall and the end of the phoneme is obtained using PRAAT. Fig. 1 also shows the F0 movement on the vowel /a/ in the word /*hadirin* / 'guest'. The F0 analysis needs only to concentrate on vowels, as for any language, vowels are the most pitch significant phones in a syllable due to more F0 movement to vowel that provide an elusive F0 contour for consonant.

Finally, F0 and duration of the recorded speech was compared with the F0 and duration of neutral speech synthesized using the existing Malay TTS system to formulate the appropriate conversion rules for each emotion. Table 3 shows the F0s extracted from the word "hadirin". Measurements of the F0 of each vowel are taken at the beginning, end, peak and fall. As the figure indicates, there are a number of patterns for the F0 movement including gradual fall, gradual rise or a mix of fall and rise. Fig. 2 shows an overall F0 contour movement of the vowels for the whole sentence "kamil belum bersedia".

The analysis of the recorded utterances provides a general view of the way in which the prosody differs from one type of emotion to another. The data obtained in the feature extraction process are analyzed to trace the relationship between emotional speech and its prosodic properties. From the analysis it is found that vowels prosodic features differ from each other but complies with a similar pattern across different emotion. For example, the F0 of anger for all vowels is higher than other type of emotion. The F0 decreases in the order of anger > happiness > fear > sadness. Duration is the longest for sadness and shortest for anger. For intensity, the descending order follows the F0, indicating that F0 has a significant influence over the speech intensity. When all the recorded utterances were analyzed for its speech properties, a general prosodic pattern for all four emotions were obtained and summarized as shown in table 4.

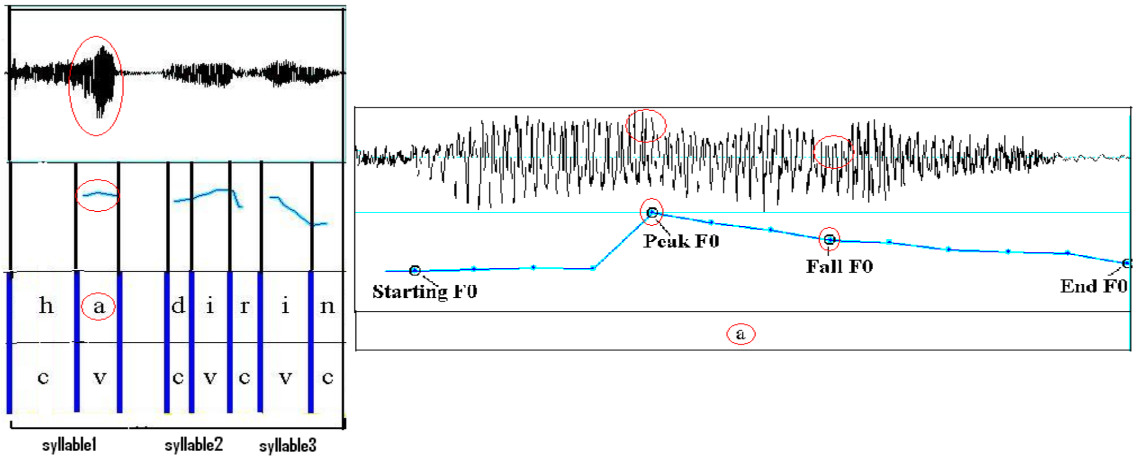
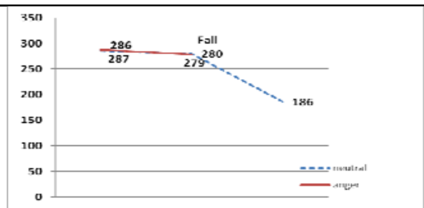
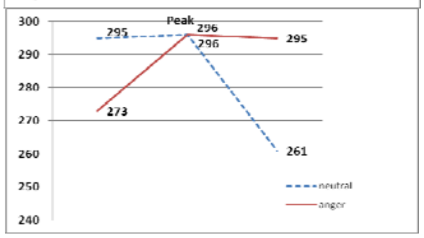
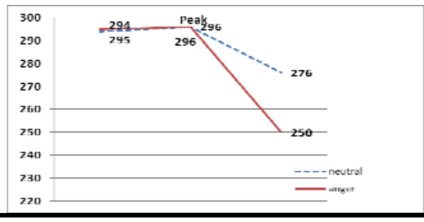


Fig.1. Segmentation and F0 extraction for vowel ‘a’ from the word “hadirin” using PRAAT

Table 3. Extracting F0 from the vowels of the word *hadirin*

Word	*Phoneme*	*F0 information*	*Neutral*	*Anger*	*F0 contour movements*
H ***a*** D ***i*** R ***i*** N	Vowel 1 (a)	Starting F0 Peak F0 Fall F0 Ending F0	286 - 280 286	287 - - 279	
	Vowel 2 (i)	Starting F0 Peak F0 Fall F0 Ending F0	295 296 - 261	273 296 - 295	
	Vowel 3 (i)	Starting F0 Peak F0 Fall F0 Ending F0	294 296 - 276	295 296 - 250	

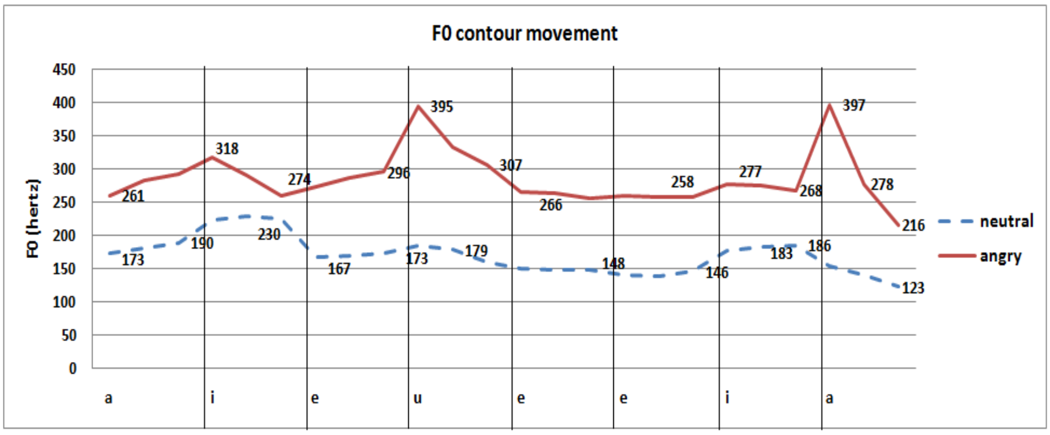


Fig.2. F0 movements in vowels in three words (kamil belum bersedia) in neutral and angry utterances

Table 4. Summary of prosodic features of Malay emotional speech

	Happy	*Angry*	*Sad*	*Fear*
Intensity	High	very high	low	Low
F0 contour	fall and rise at the beginning of sentence, falling at the end of sentence	fall at beginning of sentence, rise and fall at the end of sentence	rise and fall at the beginning of sentence, rise and fall at the end of sentence	fall at the beginning of sentence, rise and fall at the end of sentence
F0 range	wide	wide	narrow	narrow
Duration	short	shorter	long	shorter
Pauses	no pauses	short pauses	long pauses	short pauses
Others		tremor	breathiness	breathiness

4.0 SYSTEM OVERVIEW

The emotion generator developed in this research, re-synthesizes the neutral speech synthesized by an existing diphone concatenative TTS system. A diphone concatenative synthesis uses actual human recorded voice snippets as the basis for synthesis. A diphone is a unit of speech extending from the middle of one phone to the middle of the next, and in diphone synthesis, the system concatenates diphones to generate synthetic speech.

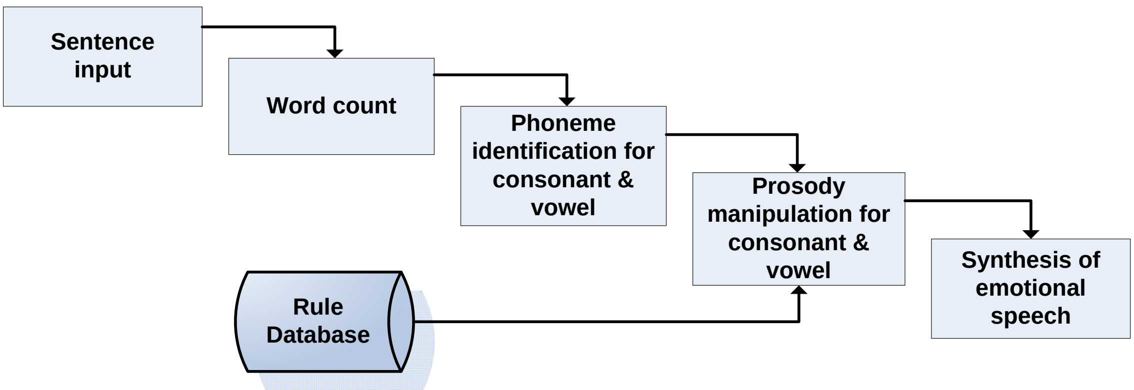


Fig.3. The process of re-synthesizing neutral speech

The initial task of the emotions generator is to determine the number of words in the target sentence in order to refer to the correct rule database. The target sentence is decomposed to its individual phoneme level and classified to consonant and vowel. Then the conversion of neutral prosody to emotional prosody takes place using the appropriate conversion rule based on the length of the sentence. For example, a two-word text input will refer to the two-word conversion rule. Intensity is manipulated manually using the MBROLA tool [21] using a common factor for the entire sentence. Finally, the emotional speech is synthesized using MBROLA tool. Figure 3 illustrates the process of re-synthesizing neutral speech with the emotions generator. The emotions generator in this research is integrated to the existing Malay speech synthesis system through the sound generation unit as depicted in Figure 4.

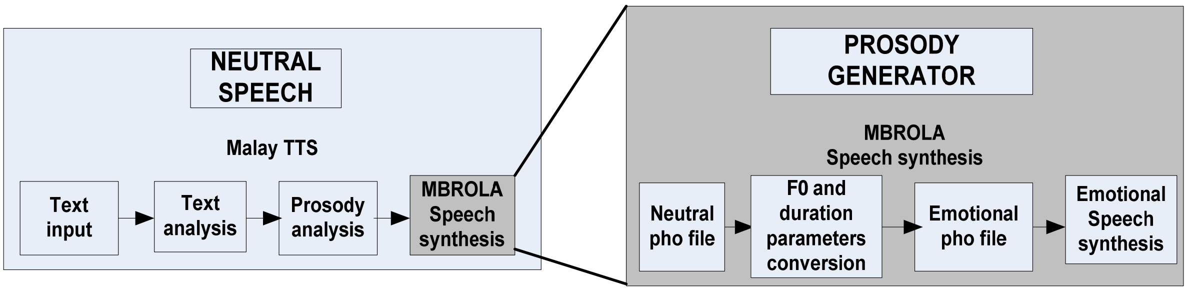


Fig.4. Overview of the prosody generator

4.1 The Prosody Conversion rules

In this research, the emotional speech is re-synthesized from neutral speech by using the prosody conversion rules established during the analysis. Each emotion has its own conversion rules in order to generate the appropriate prosody for emotional speech. The rules govern the way in which the F0, duration and intensity of neutral speech are converted. This includes the following:

(1) Increasing/decreasing the duration of a consonant
(2) Increasing/decreasing the duration of a vowel
(3) Increasing/decreasing the starting F0 of a vowel
(4) Increasing/decreasing the ending F0 of a vowel
(5) Adding pauses of different durations depending on the type of emotion
(6) Increasing/decreasing the intensity at sentence level

For example, the detailed conversion rules for anger are as follows:
Conversion rule for consonants

- Duration is modified by a factor ranging from – 62% to + 245%

Conversion rules for vowel

- Duration is modified by a factor ranging from – 41% to +132%
- Starting F0 is modified by a factor ranging from +63% to + 170%
- Ending F0 is modified by a factor ranging from + 31% to + 168%

Pauses are added with durations ranging from 80ms to 150ms depending on the length of the sentence.
Intensity is increased by 100%

5.0 TESTING AND RESULTS

Subjective evaluation of the synthesized emotional speech by the emotions generator was undertaken using a new group of listeners consisting of twenty native speakers of Malay. Similar listening tests described in section 3.3 are used in the evaluation. For this test, five new emotionally neutral sentences were constructed with different lengths and syllable structures. These sentences are presented in table 5. All five sentences were synthesized using the Malay TTS system and then re-synthesized to express four different emotions using the emotions generator. The twenty synthesized utterances (five for each emotion) were randomly shuffled and used in the listening test.

Each listener was required to listen to the twenty synthesized utterances, and then describe the emotional content according to the prosody rather than the semantic content. The test used the same open choice method of evaluation, which allows the listeners to choose one of the four emotions being evaluated or else specify any other type of perceived emotion. Table 6 summarizes the results of the test, which shows *inter alia* that happiness has the highest recognition rate at 85% and also perceived as exciting.. Anger has a recognition rate of 82% and was confused with happiness and disgust. Sadness and fear have lower recognition rates of 79% and 61% respectively. The listening test also shows that these two emotions are commonly confused with each other. In addition, sadness is confused with worry, and fear is confused with neutral speech.

Table 5. Example of emotionally neutral sentences used for listening test

Sentences	*Translation*
Sila datang lagi untuk beli-belah	*Please come again for shopping*
Selamat pagi cikgu	*Good morning teacher*
Makan malam	*Dinner*
Anggun sungguh Lina pada malam itu	*Lina look wonderful that night*
Mereka akan bertolak esok	*They will leave tomorrow*

Table 6. The confusion matrix of synthesized utterances with prosody effects

Emotion	Perceived as (%)								
	Happy	Angry	Fear	Sad	Neutral	Exited	Worry	Bored	Disgust
Happy	**85.0**	6.0	1.0	2.0	3.0	2.0	0.0	0.0	1.0
Angry	11.0	**82.0**	1.0	0.0	1.0	0.0	0.0	0.0	5.0
Fear	0.0	0.0	**61.0**	27.0	8.0	0.0	2.0	2.0	0.0
Sad	0.0	0.0	18.0	**79.0**	0.0	0.0	3.0	0.0	0.0

6.0 DISCUSSION

The results of the listening test for synthesized emotional speech differ from the earlier listening test for recorded utterances. First, happiness has the highest recognition rate at 85% followed by anger at 82%. This result was different from the earlier test in which anger had the highest recognition rate of 87.9%. These inconsistent results arise in part because anger involves other factors such as the effect of tremor that were not captured and manipulated by the emotions generator. It was also noted that anger was most often confused with happiness, and this is because the prosodic properties of happiness and anger were similar, in that both emotions involve wide F0, intensity and faster tempo. Secondly, sadness has a much higher recognition rate than fear in this perception test. This result was significantly different from the earlier test in which sadness had the lowest recognition rate of all four emotions. Another notable result shown by the second listening test is the close proximity of sadness and fear and their consequent mutual confusion. The results shows that 18% of sadness cases are misinterpreted as fear and that 27% of cases of synthesized fear are confused with sadness. A possible reason for the confusion is the similarity of the prosodic features associated with the two emotions, including lower energy and wide F0 range.

A similar study conducted by Murray and Arnott for English in [2] showed that sadness had the highest recognition rate in the case of emotionally neutral text re-synthesized to express the emotion. They also indicated that some emotions were confused with others, for example grief and disgust were confused with sadness.

7.0 CONCLUSIONS AND SUGGESTIONS FOR FUTURE RESEARCH

This research has shown that diphone synthesis can be effectively express emotion by manipulating the F0, duration and intensity. The emotions generator used in this research has the ability to control the prosodic features of neutral speech and generate emotional speech. The perception tests confirmed not only the role of F0 and duration in the expression of emotion, but also the ability of the emotions generator to compute appropriate F0s and durations to express specific emotions.

Our future work will involve investigating learning-based synthesis systems, such as HMM-based speech synthesis, and its ability to express a variety of emotions. HMM-based speech synthesis systems such as HTS are able to produce intelligible high quality speech with a natural-sounding voice. HMM-based speech synthesis also enables the manipulation of other features of speech, including intensity, voice quality and articulation. The powerful statistical tool built into HMM-based speech synthesis will also assist us in designing the most appropriate manipulation model to incorporate not only F0 and duration, but also intensity and voice quality.

8.0 ACKNOWLEDGEMENTS

This work is supported by an UMRG grant from University of Malaya, Malaysia

REFERENCES

[1] J.E. Cahn, “The Generation of Affect in Synthesized Speech”, *Journal of the American Voice I/O Society*, volume 8, 1990, pp. 1-19.

[2] I.R. Murray and J.L. Arnott, “Applying an analysis of acted vocal emotion to improve the simulation of speech synthesis,” *Journal of Computer Speech and Language 22*, 2008, pp. 107-129.

[3] A. Iida, S. Iga, N. Campbell, F. Higuchi, and M. Yasumura, “A corpus-based speech synthesis system with emotion,” *Journal of Speech Communication*, vol. 40, no. 1-2, 2003, pp. 161-187.

[4] F. Burkhardt, “Emofilt: The Simulation of Emotional Speech by Prosody-Transformation”, *in Proc. of Interspeech, 9th European Conference on Speech Communication and Technology,* Lisboan, 2005.

[5] B. Mumtaz, R.N. Ainon, M.D. Zuraidah, and G. Knowles, “Integrating Rule and Template-based Approaches for Emotional Malay Speech Synthesis”, in *Proc. of Interspeech 2008*, pp. 253-256.

[6] L.L. Syaheerah, J.M. Montero, R.N. Ainon, and M.D. Zuraidah, “eXTRA: A Culturally Enriched Malay Text to Speech System", in *Symposium on Affective Language in Human and Machine,* Aberdeen, UK, (1-2), 2008.

[7] A.A. Razak, M.I.Z. Abidin and R. Komiya, “Emotion pitch variation analysis in Malay and English voice samples”, *in the 9th Asia Pacific Conference on Communications,* 2003.

[8] T. Yoshimura, K. Tokuda, T. Masuko, T. Kobayashi, and T. Katmura, “Simultaneous modeling of spectrum, pitch and duration in HMM-based speech synthesis”, in *Eurospeech 1999,* pp. 2347-2350.

[9] A.W. Black, Z. Heiga and K. Tokuda, “Statistical parametric speech synthesis”, in *Proc. of International Conference on Acoustic Speech and Signal Processing (ICASSP)*, 2007, pp. 1229-1232.

[10] A. Black, “Unit Selection and Emotional Speech”, *Eurospeech 2003*, Geneva, Switzerland, 2003.

[11] M. Tachibana, J. Yamagashi, T. Masuko, and T. Kobayashi, “Speech Synthesis with Various Emotional Expression and Speaking Style by Style Interpolation and Morphing”, *IEICE Transaction of Information and System*, vol. E88-D, No 11, 2005, pp 2484-2491.

[12] T. Banziger and K.R. Scheree, “The role of intonation in emotional expressions”, Journal of Speech Communication, vol. 46, 2005, pp. 252-267.

[13] J.M. Montero, J. Gutiérrez-arriola, J. Colás, E. Enríquez, and J.M. Pardo, “Analysis and modelling of emotional speech in Spanish”, in Proceeding of Eurospeech ’99, Budapest, Hungary, September 1999, pp. 957-960.

[14] Y. EL-IMAN and M.D Zuraidah, “Rules and algorithms for phonetic transcription of Standard Malay”, *IEICE Transaction of Information and System*, vol. E88-D, No. 10, October 2005, pp. 2354-2372.

[15] M.O. Farid, Aspects *of Malay Phonology and Morphology-Generative approach*, Bangi, University Kebangsaan Malaysia, Press, 1980.

[16] M. Yunus Maris, *The Malay Sound System,* Kuala Lumpur, Fajar Bakti Sdn. Bhd. Malaysia, 1980.

[17] G.O. Knowles, and M.D. Zuraidah, *Word Class in Malay*, first edition, Kuala Lumpur, Malaysia, Dewan Bahasa and Pustaka, 2006.

[18] E. Navas, I. Hernaez and I. Leungo, “An Objective and Subjective Study of the Role of Semantic and Prosodic Features in Building Corpora for Emotional TTS”, *IEEE Transaction on Audio, Speech and Language Processing*, vol. 14, 2006, pp. 1117-1127.

[19] C.F. Huang, and M. Akagi, “A three-layered model for expressive speech perception”, *Journal of Speech Communication*, vol. 50, 2008, pp. 810-828.

[20] P. Boersma, and D. Weenink, 2004. Praat, doing phonetics by computer, http://www.fon.hum.uva.nl/praat/. Accessed and downloaded Jan 2006.

[21] MBROLA project homepage at http://tcts.fpms.ac.be/synthesis/mbrola.html. Accessed and downloaded on May, 2006.

BIOGRAPHY

MUMTAZ BEGUM MUSTAFA received the B.Sc. degree in Software Engineering from University of Putra Malaysia (2001) and Masters Degree in Software Engineering from University of Malaya in 2007. Currently, she is a PhD student at the Faculty of Computer Science and Information Technology, University of Malaya. Her research areas include emotional text- to-speech synthesis and speech prosody.

RAJA NOOR AINON is an Associate Professor at the Faculty of Computer Science and Information Technology, University of Malaya. Her research areas include Malay Text-to-Speech Synthesis, Multilingual Speech Recognition, Genetic Algorithms and Soft Computing.

ROZIATI ZAINUDDIN is a Professor at the Faculty of Computer Science and Information Technology, University of Malaya. Her areas of interest are in intelligent multimedia, image and speech processing, computational fluid dynamics, bio-medical informatics, computer vision visualisation and E-Learning.

ZURAIDAH MOHD DON is Professor at Faculty of Languages and Linguistics, University of Malaya. Her research interests include Prosody and Corpus Linguistics. Her recent articles have been published in Journal of Pragmatics, International Journal of the Sociology of Language, Linguistics Journal, Multilingua and International Journal of Corpus Linguistics.

GERRY KNOWLES spent 2002-2003 as Visiting Professor in the Faculty of Languages and Linguistics, University of Malaya on sabbatical from Lancaster University in the UK. His research interests include corpus linguistics, phonetics and history of English.

SALIMAH MOKHTAR is an Associate Professor at the Faculty of Computer Science & Information Technology, University Malaya. Her research interests include information system planning, information system development and evaluation, application of information technology in education and human activity system, knowledge management and computational linguistics.