Scientific Research and Essays Vol. 6(20), pp. 4281-4289, 19 September, 2011
Available online at http://www.academicjournals.org/SRE
DOI: 10.5897/SRE11.931
ISSN 1992-2248 ©2011 Academic Journals

Full Length Research Paper

Unit selection for Malay text-to-speech system using segmental context and simulated annealing

Tian-Swee Tan[1]*, Lim Yee Chea[1], Zaitul Marlizawati Zainuddin[3], Sheikh Hussain Shaikh Salleh[2] and Hum Yan Chai[2]

[1]Center for Biomedical Engineering, Transport Research Alliance, Universiti Teknologi Malaysia, 81310 UTM Skudai, Johor DT, Malaysia.
[2]Center for Biomedical Engineering, Faculty of Biomedical Engineering and Health Science, Universiti Teknologi Malaysia, 81310 UTM Skudai, Johor DT, Malaysia.
[3]Mathematics Department, Faculty of Science, Universiti Teknologi Malaysia, 81310 UTM Skudai, Johor DT, Malaysia.

Accepted 16 August, 2011

Unit selection method has become the main approach in speech synthesis. The increasing size of recorded speech has resulted in better synthesis speech quality but at the same time also resulted in more expensive computational effort. Therefore, this paper proposes a combination of segmental context matching procedure and Simulated Annealing (SA) in unit selection to improve the quality of synthetic speech and reduce the computational time. The process of unit selection is based on minimization of two costs: target cost and join cost. The segmental context (target cost), the first stage of unit selection matching procedure used to narrow down the search space, followed by an optimization method which is SA to find the units sequence with minimum join cost. Result shows that the synthesis words produced by the proposed system are 15.48% better compared to previous version of corpus-based Malay Text-to-Speech system. Future works may focus on combining SA with other heuristic methods to further enhancing the performance of unit selection.

Key words: Speech concatenation, unit selection, corpus based, heuristic method, simulated annealing.

INTRODUCTION

Since the launching of multimedia super corridor (MSC) project in Malaysia, the information and communication technology (ICT) has been growing rapidly. As a result, computer system as a tool for information and communication medium is becoming more important since then. In addition, the human computer interaction system which involved speech recognition, synthesis etc. also experiences tremendous growth resulting in many applications being developed and commercialized. For instance, Microsoft recently launched the Office XP that has the capability to pronounce (or read aloud) the text input using the speech synthesis engine. Indeed, speech synthesis has been very useful in helping human in various areas such as telephone speech, application in cars, public information systems, education assistance tools, robotic, email reading etc (Mangold, 2001; Salleh et al., 2002). The text to speech (TTS) system is also useful for the people that are physically handicap. For example, speech synthesis has been used as reading and communication tools for visually impaired patients. The first commercial TTS system is Kurzweil Reading Machine for the blind introduced by Raymond Kurzweil in the late 1970's (Klatt, 1987). For the hearing impaired and vocally handicapped, the TTS system has been used as a communication tool with people who are sign language illiterate (Tan et al., 2007a; Tan et al., 2007b; Gold and Morgan, 2000). Another application of the TTS system is helpful automatic machine for language and emotional talk (HAMLET) which is developed to help users to express their feelings (Tan et al., 2008d; Lemmetty, 2001). Speech synthesis or text to speech is

*Corresponding author. E-mail: tantswee@utm.my. Tel: +60127428412. Fax: +6075535430.

the process of transforming plain text into computer generated synthetic speech (Hasim et al., 2006). There are two types of speech synthesis methods which are parameter synthesis and concatenative synthesis (Hirai and Tenpaku, 2004). The concatenative approach is based on the idea of re-combining natural prosodic contours and phoneme sequences using a superpositional framework (Jan et al., 2005).

Corpus-based concatenative synthesis has become the major trend recently because of its highly natural speech quality (Tan and Sheikh, 2008a, 2008b; Sakai et al., 2008). Unit selection is the main component in text to speech synthesis system. It produces highly intelligible, near natural synthetic speech (Tan and Sheikh, 2008c; Tsiakoulis et al., 2008). This method creates speech by re-sequencing pre-recorded speech units selected from a very large speech database (Cepko et al., 2008). Synthetic speech through unit selection is produced by searching through large speech database (corpus) and concatenating selected units, thus forming the output signal. The selection of speech units from the database is based on minimization of target cost and concatenation cost (Clark et al., 2007; Díaz and Banga, 2006; Hunt and Black, 1996). This approach shows its superiority over formant and articulatory synthesis, because it tends to concatenate natural acoustic units with no modification. Thus, offering better speech quality (Janicki et al., 2008). However, large database can also mean costly in terms of database collection, search requirements, and segment memory storage and organization (Chappell and Hansen, 2002). Thus, a robust unit selection is needed to handle the huge volume of data in the database (Blouin et al., 2002). Viterbi algorithms (Blouin et al., 2002; Clark et al., 2007; Cepko et al., 2008; Sakai et al., 2008) is commonly used to solve unit selection problem. However, viterbi search required high computational time (Sakai et al., 2008). The heavy computational time will cause the slow generation of synthetic speech. Simulated annealing (SA) algorithm is able to solve the issue of high computational time by adjusting its control parameters.

The limitation of SA compare to commonly used viterbi search is degraded in speech quality when SA consumes less searching time. The main objective in this research is to investigate the quality of synthetic speech using SA compare to previous synthesis system proposed by Tan and Sheikh (2008a).

MATERIALS AND METHODS

The simulated annealing (SA) algorithm is based on Monte-Carlo methods and it may be considered as a special form of iterative improvement (Manuel, 1997). It was Kirkpatrick et al. (1983) who first proposed SA as a method for solving combinatorial optimization problems. In general, SA algorithm applicable for solving combinatorial optimization problems by generating a sequence of moves at descending values of a control parameter (Jeong and Kim, 1990). The aim of SA is to choose a good solution to an optimization problem according to some cost function on the state space of possible solutions (Rose et al., 1990). SA is a generalization of the local search algorithm. In the iterative process for SA algorithm, its algorithm allowed accepting non-improving neighboring solutions (Turgut et al., 2003) to avoid being trapped at a poor local optimum with a certain probability, whereas other iterative improvement algorithm would allow only cost-decreasing ones to be accepted.

Procedure of unit selection

Unit selection starts with segmental context matching process. The selection of appropriate speech phoneme unit takes into consideration of match of left and right segmental context or textual content for each of the required phoneme. For example, the synthesis word “nasi” required phonemes /n/, /a/, /s/ and /i/. The required left and right segmental contexts for phoneme /s/ are /a/ and /i/ respectively. When unit selection is searching for candidate units for phoneme /s/, it will retain the unit with the desired surrounding segmental context only. This is called segmental context target cost minimization and is the first step of unit selection. Next, the retain candidate units will go through join cost minimization. First, an initial solution needs to be generated. In the beginning, the initial configurations including initial temperature and annealing schedule need to be determined. After the initial temperature is chosen, generate an initial solution and its cost function value. The initial temperature is set at a high level so that almost all moves will be accepted initially. This initial solution is defined as current solution. Next, obtain a neighboring solution of the current solution using local search technique. The temperature is lowered according to annealing schedule along the algorithm until almost no moves will be accepted. Obtain the cost of the neighboring solution and compare it to the cost of the current solution. If the cost is better, it will be accepted as current solution. Else, the new solution may be accepted as current solution only when the Metropolis's criterion is met which is based on Boltzmann’s probability. This process then continues from the new current solution and SA algorithm stops when stopping criteria is met. Figure 1 shows SA flow diagram to find best speech unit sequence. According to Metropolis's criterion (Metropolis et al., 1953) as shown in Figure 2, if the new cost is lower than the current solution in the case of minimization, it is updated as current solution. Else, it is accepted with probability P where P is Boltzmann’s equation.

Probability of accepting a non-improving solution as the current solution is based on Boltzmann’s equation represented by Equation 1, where T is the temperature; ΔE is different between new cost and current solution. Next, a random number λ in (0, 1) is generated from a uniform distribution and compare it with value P. If $P > \lambda$, then the new solution is accepted as current solution. Else, it is rejected.

$$P(\Delta E) = \exp\left(-\frac{\Delta E}{T}\right) \quad (1)$$

Computational of concatenation cost

To measure the spectral distance between two join segments, the final frame’s mel frequency cepstral coefficients (12 coefficients) of the current unit will pair with initial frame’s mel frequency cepstral coefficients (12 coefficients) of the next unit. Then the 12 coefficients of initial and final frame of speech unit were used for distance measure. The Euclidean distance will then take the 12 coefficients of initial and final frame to calculate spectral distance between two joined segments. We called this as local concatenation cost or join cost. To calculate concatenation cost in a

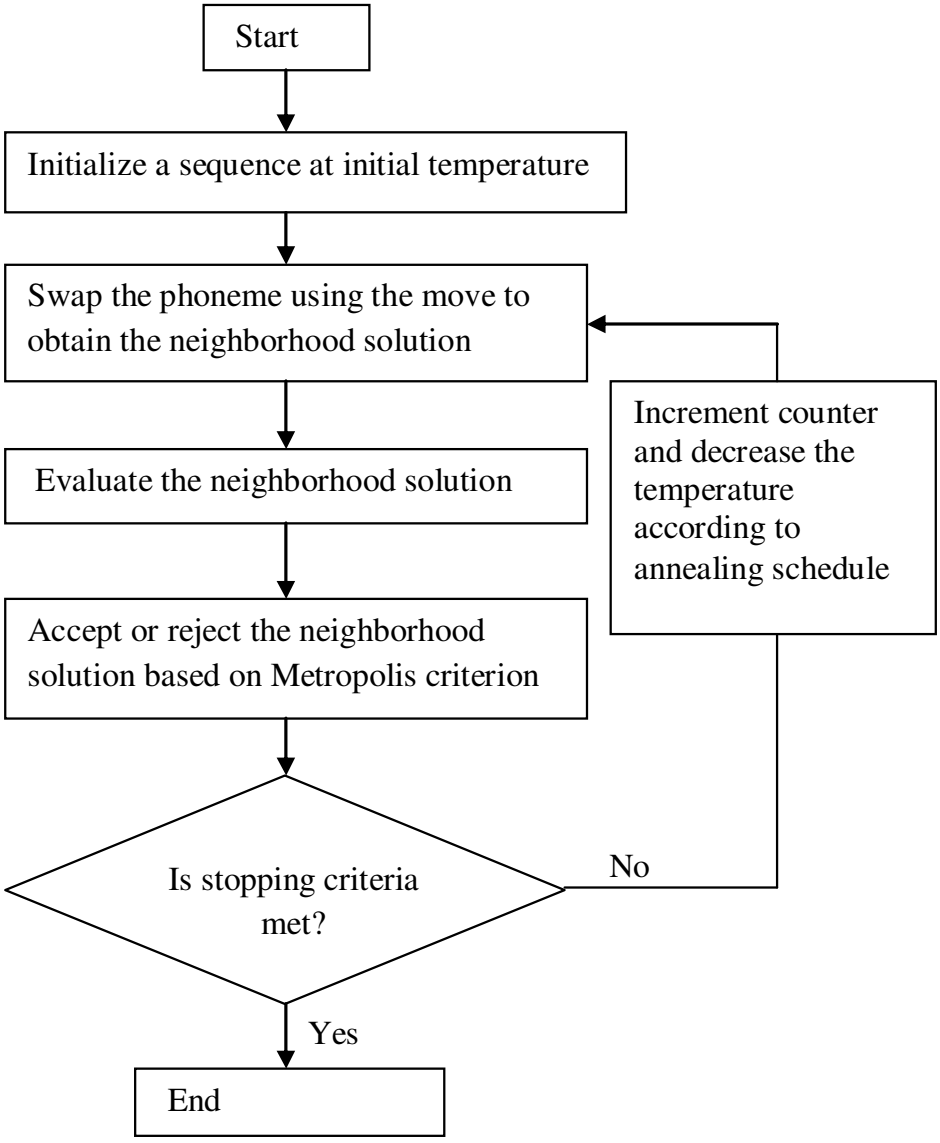

Figure 1. Simulated annealing flow diagram to find best speech unit sequence.

word, it is simply total up the local concatenation cost in a word. SA algorithm plays the role of pairing and exploring various combinations of unit sequences in concatenation cost minimization. Initial solution is needed to initialize SA algorithm. In this research, the initial solution is fixed for all the test problems. The first candidate of the correspondence phoneme after segmental context matching process is chosen as an initial solution. The objective function for concatenation cost minimization in unit selection is represented by:

$$u^* = \arg \min_{u \in U} \left(\sum_{i=1}^{N} C_C \left(u_i, u_{i-1} \right) \right) \quad (2)$$

Where, $u = u_1, u_2, \ldots u_N$ denote the units in the inventory U which minimize the concatenation cost in Equation 2. Equation 2 is used as spectral distance measurements and also as a quantitative measurement of quality of synthetic speech. To get the local concatenation cost value, it requires the parameterization of units and a distance measure. Concatenation cost between units u_i and u_{i-1} can be written as:

$$C_c \left(u_i, u_{i-1} \right) = \sqrt{\sum_{j=1}^{12} \left(u_i - u_{i-1} \right)^2}$$

The choice of a cooling schedule has an important effect on performance of SA algorithm (Ali et al., 2002). For this reason, modifications and improvements have been tried by tuning the parameters (cooling rate) for better quality or time tradeoff. These temperature values are controlled by a cooling schedule that specifies the initial and decreasing temperature values at each stage of the algorithm. The following geometric function has been

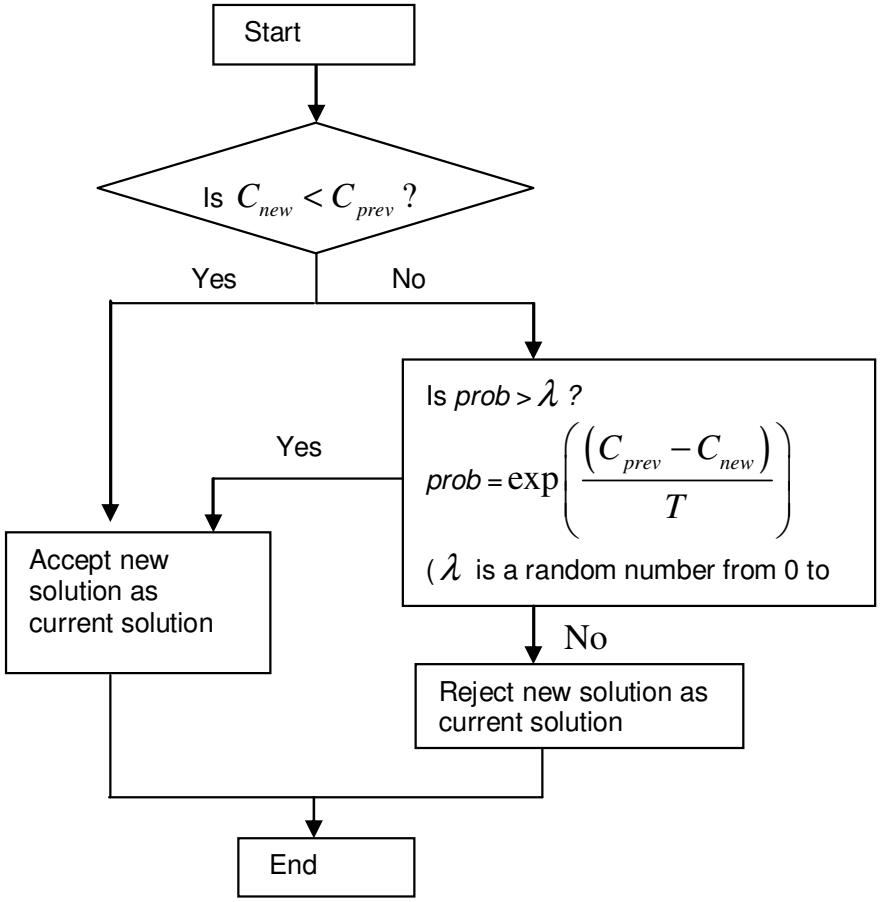

Figure 2. Metropolis criterion (McGookin and Murray-Smith, 2006).

taken as the temperature reduction function:

$$T_{k+1} = \alpha T_k \quad k = 0, 1, 2, 3,.. \quad 0 < \alpha < 1$$

Where T_k is the temperature at stage *k*, α is the temperature reduction rate. In this research, the various temperature reduction rates was tested which are 0.80, 0.85, 0.90 and 0.95. Figure 3 shows the temperature reduction pattern for these temperature reduction rates for length of Markov chain equal to one. The initial temperature T_0 is set relatively high so that most of the moves are accepted in the early stages and there is little chance of the algorithm intensifies into the region of local minimum. The initial temperature and final temperature (Chen and Su, 2002) in unit selection is set according to the Equations 3 and 4 respectively,

$$T_i = \frac{-1}{\ln P_i} \tag{3}$$

$$T_f = \frac{-1}{\ln P_f} \tag{4}$$

Where P_i is the desired initial probability and P_f is the desired final probability.

The parameters values in unit selection are given as follows:

i) Initial temperature, $T_0 = \frac{-1}{\ln 0.999} = 999.499 \approx 1000$

ii) Final temperature, $T_f = \frac{-1}{\ln 0.00001} = 0.0869 \approx 0.1$

iii) Temperature reduction rate, $\alpha = 0.80, 0.85, 0.90, 0.95$.

Stopping criteria used is algorithm stop when number of non-improving iterations of 200 is reached or final temperature $T_f < 0.1$. The length of the Markov chain is required to decide how many trials are to be used at each value of T. Markov chain used here was reduced by the temperature according to annealing schedule after two successive iterations. Markov chain length is simply fixed to 2 since this is not main focus of this research.

Neighborhood generation mechanism

The neighborhood generation mechanism apply here is randomly swap a phoneme per iteration. Example:

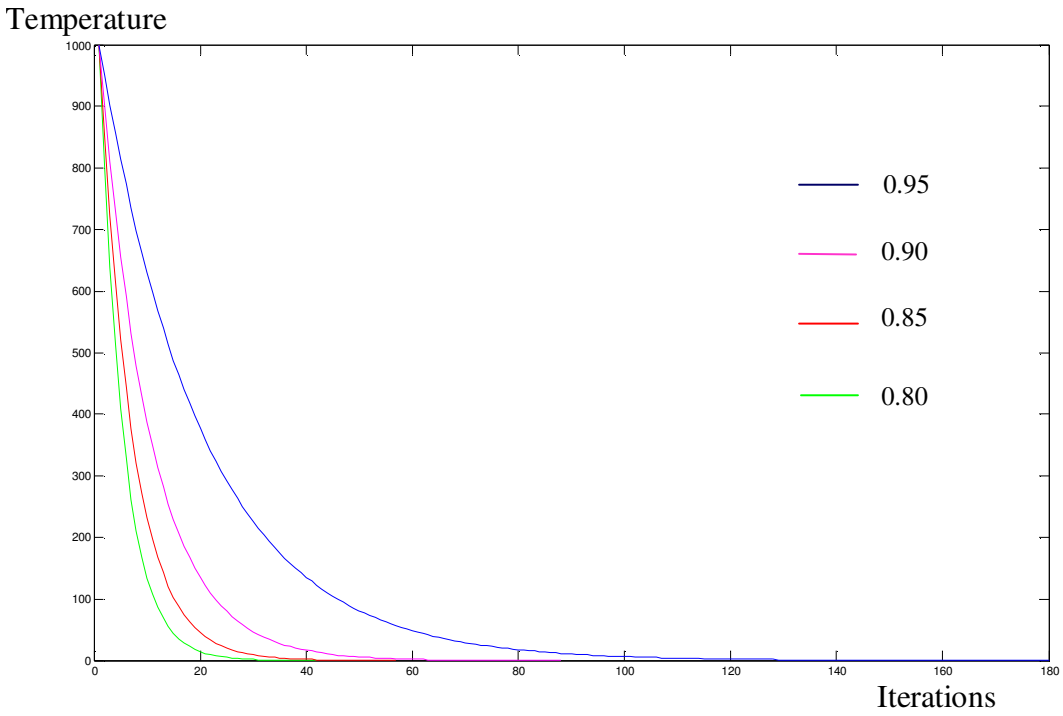

Figure 3. Temperature reduction pattern for various reduction rates with Markov chain length 1.

Initial solution

The number in bracket, "[ ]" represents the units candidate number after matching process (Figure 4). The join cost is computed as:

$$C_c\left(u_i, u_{i-1}\right) = \sum_{i=1}^{i=n} Local\ \ Cost\left[i\right] = \sum_{i=1}^{i=n} \sqrt{\sum_{j=1}^{12} \left(u_i - u_{i-1}\right)^2}$$

Where *n* depicts the total number of local cost. Local cost here refers to only concatenation cost.

Iteration 1

Apply move 1 to obtain neighborhood solution.

Neighborhood solution

The phoneme 3 is chosen randomly to swap. The candidate's number is changed randomly from 1 to 7 for example. When phoneme 3 is changed, the local cost (*i* + 1) and local cost (*i* + 2) will be changed while local cost (*i*) and local cost (*i* + 3) will remain unchanged (Figure 5). If this neighborhood solution is accepted as current solution, then it generates another neighborhood solution based on this newly accepted solution.

Iteration 2

Neighborhood solution

The phoneme 1 is chosen randomly to swap. The candidate's number is changed randomly from 1 to 9 for example. When phoneme 1 is changed, only the local cost (*i*) will be changed while other local cost will remain unchanged (Figure 6). If this neighborhood solution is accepted as current solution, then generate another neighborhood solution based on this newly accepted current solution.

Iteration 3

Neighborhood solution

The phoneme 4 is chosen randomly to swap. The candidate's number is changed randomly from 1 to 8. When phoneme 4 is changed, the local cost (*i* + 2) and local cost (*i* + 3) will be changed while local cost (*i*) and local cost (*i* + 1) remain unchanged (Figure 7). If this neighborhood solution is rejected as current solution, then go back to the previous iterations and generate another neighborhood solution based on the solution of previous iterations (current solution).

RESULTS AND DISCUSSION

To evaluate the performance of the proposed system and previous version of corpus-based Malay text-to-speech system, comparison of join cost in unit selection which correspondence to perceptual scores is conducted. The proposed system in this research is actually exactly the same as previous version except the unit selection, computational of join cost minimization. Therefore, join cost comparison is able to distinguish the performance between two systems since everything is the same

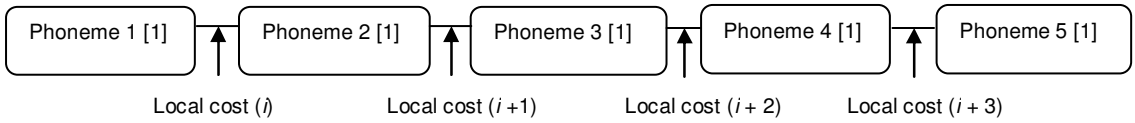

Figure 4. Initial solution.

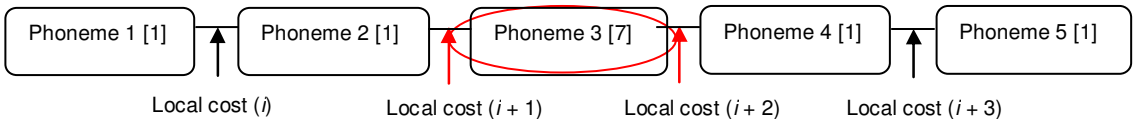

Figure 5. Iteration 1 neighborhood solution.

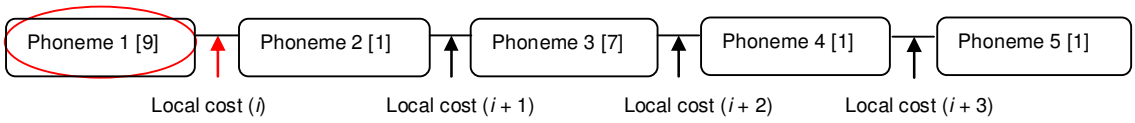

Figure 6. Iteration 2 neighborhood solution.

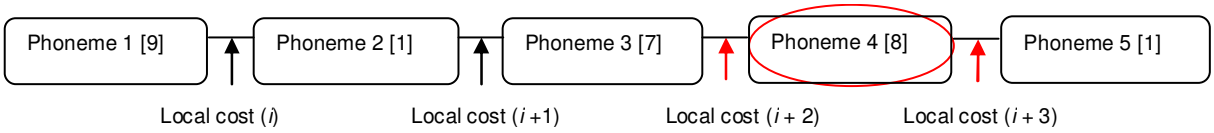

Figure 7. Iteration 3 neighborhood solution.

except the join cost minimization. Table 1 shows the 40 Malay words selected for join cost comparison that covers almost all Malay phoneme set using Equation 2. Join cost I represents join cost obtained by the previous version of synthesis system while Join cost II represents mean join cost obtained by the proposed system using SA algorithm. Smaller value of join cost indicates the smaller spectral discontinuity at the join segments and better quality of synthetic speech. The join cost obtained for 38 words produced by the proposed system is better than the previous version with an average improvement of 15.48% for 40 words. Thus, this indicates that the synthesis words produced by the proposed system have better smoothness of join boundary. Table 2 shows the distribution of 40 words in terms of different magnitude of improvement in join cost. The class which has the highest frequency is range between 20 to 30%. SA consumed 1448 milliseconds in obtaining the optimum solution of 40 words in Table 1 while previous system consumes 1391 milliseconds whereas Viterbi search consumes 7285 milliseconds. This set of computational time are obtained using a set of computers PC Intel Core 2 Duo 1.58 GHz, 2.00 GHz of RAM. SA consumes 4.17% extra computational time than previous system but resulted in 15.48% better synthetic speech quality than previous system.

A formal listening test involved 40 Universiti Teknologi Malaysia students with no hearing loss was conducted to evaluate the output sound. 55% of the listeners are female and the rest are male listeners. The ages of listeners were range between 21 and 27, with a mean age of 24 years old. 35% of the listeners were native speakers of Malay language while the rest were not. The listening test was conducted individually using headphone and a set of computers PC Pentium IV 3 GHz in a quiet room. The testing methods that can be used in corpus-based Malay text-to-speech system were modified rhythm test (MRT), mean opinion score (MOS) and perceptual test which is also known as intelligibility test (Rilliard and Aubergé, 2001). The ‘modified rhyhm test’ and ‘mean opinion score’ can be group in ‘auditory test’. Modify rhythm test, word spotting test and mean opinion score listening test were conducted. The aims of ‘modify rhythm test’ are to evaluate the accuracy and verify the intelligibility of the synthetic speech sound especially in pronunciation. A set of 50 questions with

Table 1. Words selected for join cost comparison.

Words	No. of phoneme	Join cost I	Join cost II	Improvement (%)
Nasi	4	44.92	38.87	13.47
Musim	5	57.64	55.84	3.12
Janji	5	65.24	55.86	14.38
Kampung	6	51.65	43.35	16.07
Vitamin	7	68.41	56.02	18.11
Demikian	7	70.19	58.38	16.83
Muktamad	8	61.79	53.43	13.53
Informasi	9	93.66	90.18	3.72
Selanjutnya	10	117.73	88.63	24.72
Berpengetahuan	12	113.45	91.68	19.19
Siapa	4	32.36	30.48	5.81
Adik	4	33.43	25.98	22.29
Bahasa	6	80.23	59.63	25.68
Banyak	5	32.34	30.09	6.96
Tanah	5	43.20	39.95	7.52
Lampu	5	33.51	34.75	-3.70
Keluarga	7	69.67	51.66	25.85
Jadual	5	45.19	37.90	16.13
Istana	6	65.12	52.03	20.10
Tanggungjawab	11	86.34	81.38	5.74
Cahaya	6	58.68	42.89	26.91
Daun	3	19.51	17.31	11.28
Tuan	3	27.23	25.03	8.08
Utama	5	37.49	36.21	3.41
Kunci	5	44.07	40.36	8.42
Lazim	5	41.72	46.94	-12.51
Nasional	7	78.94	62.58	20.72
Olahraga	8	68.68	60.39	12.07
Pakai	4	37.10	33.70	9.16
Faham	5	48.90	29.38	39.9
Disyaki	6	75.19	60.75	19.20
Khidmat	6	55.89	44.54	20.31
Hospital	8	97.66	72.16	26.11
Pampas	6	63.77	59.84	6.16
Raksasa	7	96.49	75.84	21.40
Syarikat	7	65.07	57.79	11.19
Walaupun	7	61.04	46.53	23.77
Gelincir	8	96.15	68.89	28.35
Ialah	4	21.87	16.66	23.82
Nyawa	4	22.41	14.37	35.88

very similar phonetics sound have been tested. The aim of 'word spotting test' is to verify whether the pronunciation of letter to sound rule of synthesis system is intelligible enough for the listeners to catch it. Listeners are required to write down the answer from what they have heard from the listening test. Ten Malay words with different length have been tested. The naturalness, pronunciation and intelligibility of the output sound were tested through mean opinion score listening test.

Ten Malay words have been tested for 40 participants. Result of listening test for previous version and proposed unit selection method is presented in Table 3.

Conclusion

This paper is a first version of unit selection using 'simulated annealing' for corpus-based Malay text-to-

Table 2. Distribution of 40 words in terms of different magnitude of improvement in join cost.

Category	Frequency
Word ≤ 0%	2
0% < word ≤ 10%	11
10% < word ≤ 20%	12
20% < word ≤ 30%	13
Word > 30%	2
Total	40

Table 3 Result of the listening test.

	Previous version	Proposed system
Word spotting test	4 words: 40/40	4 words: 40/40
Score: (*x*/40)	3 words: 38/40	2 word : 39/40
*Perfect score: (40/40)	2 words: 36/40	2 words: 38/40
	1 word : 34/40	1 words: 36/40
		1 words: 35/40
Modify rhythm test	86.66%	88.54%
Mean opinion score test		
Naturalness	3.9	4.1
Pronunciation	4.1	4.2
Intelligibility	3.9	4.1
*Perfect score: (5.0/5.0)		

speech system. This system has achieved its aim of improving the speech quality compare to previous version of corpus-based Malay text-to-speech system. The unit selection is based on two cost functions which are target cost and concatenation cost. In the proposed method, the segmental context is used as a target cost for matching process. The retain candidate units are used as an input for SA in concatenation cost minimization. The listening test and values of join cost obtained have justified the improvement of speech quality of the proposed system. Therefore, SA is a suitable method for unit selection since it has made a contribution in improving the speech quality by selecting the best speech unit sequence within reasonable computational time. For future research, HMM-based speech synthesis (Zen et al., 2009; Pucher et al., 2010; Tokuda et al., 1995, Chomphan and Kobayashi, 2008) can also be developed for Malay language since it has gained attention of many researchers recently due to its flexibility in generating speech from parameter generation algorithm. Since the performance of 'simulated annealing' depends highly on parameters setting, therefore, parameters tuning is one possible approach in future research to improve the performance of 'simulated annealing'. Other heuristic method such as 'genetic algorithm' and Tabu search can also be conducted in unit selection.

ACKNOWLEDGEMENT

This research project is supported by Universiti Teknologi Malaysia's university research grant (GUP-Tier 1, Q.J1300000.7136.01H49).

REFERENCES

Ali MM, Törn A, Viitanen S (2002). A direct search variant of the simulated annealing algorithm for optimization involving continuous variables. Comput. Operations Res., 29(1): 87-102.

Blouin C, Rosec O, Bagshaw PC, Alessandro C (2002). Concatenation cost calculation and optimisation for unit selection in TTS. Proceedings of 2002 IEEE Workshop on Speech Synthesis, pp. 11-13 September. Santa Monica, USA. pp. 231-234.

Cepko J, Talafova R, Vrabec J (2008). Indexing join costs for faster unit selection synthesis. Systems, Signals and Image Processing, IWSSIP 2008. 15th International Conference.25-28 June. Bratislava, Slovakia: pp. 503-506.

Chappell DT, Hansen JHL (2002). A comparison of spectral smoothing methods for segment concatenation based speech synthesis. Speech Commun., 36(3-4): 343-373.

Chen TY, Su JJ (2002). Efficiency improvement of simulated annealing

in optimal structural designs. Adv. Eng. Software, 33(7-10): 675-680.
Chomphan S, Kobayashi T (2008). Tone correctness improvement in speaker dependent HMM-based Thai speech synthesis. Speech Commun., 50(5): 392-404.
Clark RAJ, Richmond K, King S (2007). Multisyn: Open-domain unit selection for the Festival speech synthesis system. Speech Commun., 49(4): 317-330.
Díaz FC, Banga ER (2006). A method for combining intonation modelling and speech unit selection in corpus-based speech synthesis systems. Speech Commun., 48(8): 941-956.
Hasim S, Tunga G, Yasar S (2006). A Corpus-Based Concatenative Speech Synthesis System for Turkish. Turk. J. Elect. Eng. Comput. Sci., 14(2): 209-223.
Hirai T, Tenpaku S (2004). Using 5 ms segments in concatenative speech. Fifth ISCA ITRW on Speech Synthesis. 16 Jun. Pittsburgh, PA, USA: 37-42.
Hunt A, Black A (1996). Unit selection in a concatenative speech synthesis system using large speech database. Proc. Int. Conf. Acoust., Speech, Signal Process. Atlanta, GA: pp. 373–376.
Jan VS, Alexander K, Esther K, Taniya M (2005). Synthesis of prosody using multi-level unit sequences. Speech Commun., 46(3-4): 365-375.
Janicki A, Meus P, Topczewski M (2008). Taking advantage of pronunciation variation in unit selection speech synthesis for polish. Communications, Control and Signal Processing, 2008. ISCCSP 2008. 3rd International Symposium. 12-14 March. St. Julians: pp. 1133-1137.
Jeong CS, Kim MH (1990). Fast parallel simulated annealing for traveling salesman problem. Neural Networks, 1990, IJCNN International Joint Conference. 17-21 June. Washington, D.C: pp. 947-953.
Kirkpatrick S, Gelatt CD, Vecchi MP (1983). Optimization by Simulated Annealing. Science, 220: 671-680.
Lemmetty S (1999). Review of Speech Synthesis Technology. Helsinki University of Technology: Master Thesis.
Mangold H (2001). Speech Technology in Reality - Applications, Their Challenges and Solutions. Text, Speech and Dialogue 4th International Conference, TSD 2001. September 11-13. Zelezna Ruda, Czech Republic: LNAI , 2166: 197-201.
Manuel DA (1997). Constructing efficient simulated annealing algorithms. Discrete Appl. Math., 77(2): 139-159.
Metropolis N, Rosenbluth AW, Rosenbluth MN, Teller AH, Teller E (1953). Equation of state calculation using fast computing machines. J. Chem. Phys., 21: 1087-1092.
Pucher M, Schabus D, Yamagishi J, Neubarth F, Strom V (2010). Modeling and interpolation of Austrian German and Viennese dialect in HMM-based speech synthesis. Speech Commun., 52(2): 164-179.
Rilliard A, Aubergé V (2001). Prosody evaluation as a diagnostic process: subjective vs. objective measurements. 4th Speech Synthesis Workshop. 29 August - 1 September. Scotland, ISCA: pp. 140-144.
Rose J, Klebsch W, Wolf J (1990). Temperature measurement and equilibrium dynamics of simulated annealing placements. Computer-Aided Design of Integrated Circuits and Systems, IEEE Trans. March 1990: pp. 253-259.
Sakai S, Kawahara T, Nakamura S (2008). Admissible stopping in viterbi beam search for unit selection in concatenative speech synthesis. Acoustics, Speech and Signal Processing, 2008. ICASSP 2008. IEEE International Conference. 31 March-4 April. Las Vegas, Nevada, U.S.A: pp. 4613-4616.
Tan TS, Ariff AK, Salleh SH, Siew KS, Leong SH (2007b). Wireless data gloves Malay sign language recognition system. 2007 6th International Conference on Information, Communications and Signal Processing, ICICS ISBN: 1424409837; 978-142440983-9.DOI: 10.1109/ICICS.2007.4449599.
Tan TS, Salleh SH (2008a). Corpus-based Malay text-to-speech synthesis system. APCC 2008. 14th Asia-Pacific Conference on Communications, 2008. 14-16 October: pp. 1-5.
Tan TS, Salleh SH (2008b). Implementation of Phonetic Context Variable Length Unit Selection Module for Malay Text to Speech. Science Publications. J. Comput. Sci., 4(7): 550-556.
Tan TS, Salleh SH (2008c). Corpus Design for Malay Corpus-based Speech Synthesis System. Am. J. Appl. Sci., 6(4): 696-702.
Tan TS, Salleh SH, Ariff AK, Ting CM, Siew KS, Leong SH (2007a). Malay sign language gesture recognition system 2007 International Conference on Intelligent and Advanced Systems, ICIAS 2007: pp. 982-985.
Tan TS, Salleh SH, Chew KM, Lim SC (2008d). Photo-realistic text-driven Malay talking head with multiple expression Proceedings of the International Conference on Computer and Communication Engineering 2008, ICCCE08: ISBN: 978-142441692-9. DOI: 10.1109/ICCCE.2008.4580697 Global Links for Human Development: pp. 711-715.
Tokuda K, Kobayashi T, Imai S (1995). Speech parameter generation from HMM using dynamic features. In: IEEE International Conferences on Acoustics, Speech, and Signal Processing (ICASSP*:* pp. 660–663.
Tsiakoulis P, Chalamandaris A, Karabetsos S, Raptis S (2008). A statistical method for database reduction for embedded unit selection speech synthesis. Acoustics, Speech and Signal Processing, 2008. ICASSP 2008. IEEE International Conference. 31 March-4 April. Las Vegas, Nevada, USA: pp. 4601-4604.
Turgut D, Turgut B, Elmasri R, Le TV (2003). Optimizing clustering algorithm in mobile ad hoc networks using simulated annealing. Wireless Communications and Networking, 2003. WCNC 2003. IEEE.20-20 March. New Orleans, Louisiana, USA: pp. 1492-1497.
Zen H, Tokuda K, Black A (2009). Statistical parametric speech synthesis. Speech Communication, 51(11): 1039-1064.