American Journal of Applied Sciences 6 (4): 696-702, 2009
ISSN 1546-9239

Corpus Design for Malay Corpus-based Speech Synthesis System

Tian-Swee Tan and Sh-Hussain
Faculty of Biomedical Engineering and Health Science, P11, Center for Biomedical Engineering,
University Technology Malaysia, 81310 UTM Skudai, Johor DT, Malaysia

Abstract: Problem statement: Speech corpus is one of the major components in corpus-based synthesis. The quality and coverage in speech corpus will affect the quality of synthesis speech sound. **Approach:** This study proposes a corpus design for Malay corpus-based speech synthesis system. This includes the study of design criteria in corpus-based speech synthesis, Malay corpus based database design and the concatenation engine in Malay corpus-based synthesis system. A set of 10 millions digital text corpuses for Malay language has been collected from Malay internet news. This text corpus had been analyzed using word frequency count to find out all high frequency words to be used for designing the sentences for speech corpus. **Results:** Altogether 381 sentences for speech corpus had been designed using 70% of high frequency words from 10 million text corpus. It consists of 16826 phoneme units and the total storage size is 37.6Mb. All the phone units are phonetically transcribed to preserve the phonetic context of its origin that will be used for phonetic context unit. This speech corpus had been labeled at phoneme level and used for variable length continuous phoneme based concatenation. Speech corpus is one of the major components in corpus-based synthesis. The quality and coverage in speech corpus will affect the quality of synthesized speech sound. **Conclusion/Recommendation:** This study has proposed a platform for designing speech corpus especially for Malay Text to Speech which can be further enhanced to support more coverage and higher naturalness of synthetic speech.

Key words: Text to speech, unit selection, concatenation, corpus-based speech synthesis, speech synthesis, variable length unit selection

INTRODUCTION

Nowadays, a new concept of speech synthesis: Corpus based speech synthesis (unit selection), has been introduced[1]. It has emerged as a promising methodology to solve the problems with the fixed-size unit inventory synthesis, e.g., diphone synthesis[2]. Corpus-based approaches to speech synthesis have been advocated to overcome the limitations of concatenative synthesis from a fixed acoustic unit inventory[1]. The frequency of unit concatenations in e.g., diphone synthesis has been argued to contribute to the perceived lack of naturalness of synthesis speech.

Definition and concept of corpus-based TTS: Corpus-based TTS system creates the output speech by selecting and concatenating units (e.g., Speech sounds or words) from a large (can be up to several hours long) speech database to select the units to be concatenated[2-5].

The idea of corpus-based or unit selection synthesis is that the corpus is searched for maximally long phonetic strings to match the sounds to be synthesized[4]. According to Nagy[1], as the length of the elements used in the synthesized speech increases, the number of concatenation points decreases, resulting in higher perceived quality. If the database offers sufficient prosodic and allophonic coverage, it is then even possible to generate natural sounding prosody without resort to signal manipulation.

Rutten[6] also states that with the arrival of corpus-based synthesis, segmental voice quality is further improved because this new synthesis technique aims at reducing speech modeling and manipulation in favor of speech unit selection. If the database offers sufficient prosodic and allophonic coverage, it is then even possible to generate natural sounding prosody without resort to signal manipulation. However, signal processing is still needed for segment concatenation and system flexibility (e.g., rate and pitch level changes).

Corresponding Author: Tian-Swee Tan, Faculty of Biomedical Engineering and Health Science,
Center for Biomedical Engineering, University Technology Malaysia, Malaysia
Tel: +60127428412 Fax: +6075535430

This technique is capable to search for maximally long phonetic strings to match the sounds to be synthesized[4]. It uses a large inventory to select the units to be concatenated[2].

Advantages: As compared to diphone or triphone synthesis, corpus-based speech tends to elicit considerably higher ratings of naturalness in auditory tests[4]. This is because the number of real concatenation points become much smaller[1,3].

Moreover, the database of traditional diphone and triphone concatenation TTS systems is recorded with monotonous prosody whereas the units from a large speech corpus retain their natural and varied prosody. Thus, it becomes possible to concatenate larger chunks of natural speech, providing superior quality over diphone and triphone concatenation[1-3,5].

As the corpus in its entirety provides the acoustic basis for such synthesis, the development of an optimal corpus represents an essential task of corpus-based synthesis[2,4].

The key idea of corpus-based synthesis, or unit selection, is to use an entire speech corpus as the acoustic inventory and to select at run time from this corpus the longest available strings of phonetic segments that match a sequence of target speech sounds in the utterance to be synthesized, thereby minimizing the number of concatenations and reducing the need for signal processing. In an ideal world, the target utterance would be found in its entirety in the speech database and simply played back by the system without any concatenations and without any signal processing applied, effectively rendering natural speech.

MATERIALS AND METHODS

Method and development: According to Fung and Meng[7], the development steps can be summarized as below:

- Corpus development for recording
- Waveform segmentation
- Unit selection for concatenative synthesis

According to Chou[5], a large speech corpus (on the order of 10 h, or even much more, of speech) produced by a single speaker is collected. The corpus is designed so that almost all linguistic and prosodic features for the target language (either for general domain or specific domain) have been included. Parallel analysis of all prosodic and linguistic features of the speech signals as well as the corresponding texts can lead to a much better prosodic model. There can be many repetitions of a given voice unit in the corpus, but in different context with different prosodic features.

In a different approach, longer elements (such as phrases, words or syllables) are also labeled in the speech database and can be selected directly (without the implicit selection mechanism of a cost function)[1,6].

Speech corpus design criteria: The criteria in designing speech corpus are size, corpus coverage, domain and quality as shown in Fig. 1. This section will discuss on the design criteria and conclude with the CBMTTS speech corpus' design specification.

Size of corpus: The recent developed corpus-based speech synthesizers tend to rely on large-scale database, several vary from hours to more than 10 h speech corpora to provide sufficiently natural output speech[8]. But the increase of corpus size will affect the unit selection time and slow down the process of synthesis[4].

Consequently, the size of corpus is the first issue to be tackled before other issues. This meant a comparison between the minimum and maximum sizes. A maximum size would mean a greater probability of the corpus containing the biggest possible units to match the text to be synthesized- from sound strings to words or even phrases[4]. Unfortunately, big databases have been found complicated to maintain and even more complicated to annotate[9]. Moreover, segmentation and tagging of corpus units is a cumbersome and time-consuming process - it has been found that one-minute corpus takes 1000 min to mark up. This is why some people decided to make the corpus as small as possible, yet containing as much relevant material as possible[4].

The design of a well utilized speech corpus requires determining an optimal set of elements for storage in the database. Optimality in this case means finding equilibrium between a large number of elements demanded by quality requirements and a minimal element number constrained by performance considerations[1].

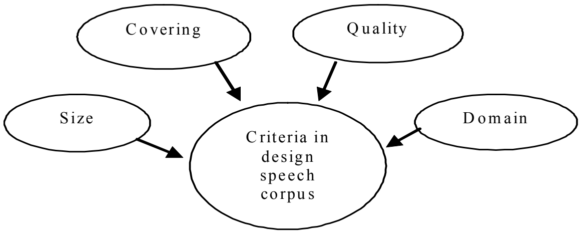


Fig. 1: Criteria in designing speech corpus

Corpus database coverage: According to Ni[10], as anyone who has built a unit selection synthesizer, knows the quality of synthetic speech is highly dependent on the unit coverage of a speech corpus.

The corpus database design shall be “phonetically rich corpus”. In other word, it is to include as many as possible phonetic combinations, including intra-syllabic and inter-syllabic structures, in a corpus of acceptable size[5]. The corpus words should also include at least one instance of all possible units[1]. For instance, for Estonian speech which has diphones as smallest units, the first database for Estonian speech synthesis consisted of ca 1700 diphones.

To ensure this, some projects split the texts into two parts, designed along different lines. The first part provides coverage of frequent words and phrases which determined from the statistical properties of the target domain and allow selecting the longest possible elements for concatenation. The other part ensures coverage of diphones for the diphone based synthesis[1].

From the Estonian text analysis, it is concluded that the 10 most frequent word forms covered 31% of the input corpus. As little as 500 words ensure 92% coverage, while with 2300 words reach 99%. A corpus from an unrestricted domain requires approximately 70,000 word forms to reach 90% coverage[1].

According to Chou[5], the corpus is designed so that almost all linguistic and prosodic features for the target language (either for general domain or specific domain) have been included. Parallel analysis of all prosodic and linguistic features of the speech signals as well as the corresponding texts can lead to a much better prosodic model. There can be many repetitions of a given voice unit in the corpus, but in different context with different prosodic features.

Corpus database quality: According to Piits[4], a system with a good selection module and a high-quality speech corpus may yield output speech of extremely high quality, even if the signal processing module is rather simple.

The final quality of a concatenation synthesis system is directly related to the continuity of the spectrum at the concatenation point. A key factor in the final quality of a text-to-speech system based on unit concatenation is the continuity, or smoothness, of the formant trajectories at the concatenation point[11].

For high quality speech synthesis, the recording script should offer variety of comprehensive word[12]. Besides that, it should also cover all the prosodic and acoustic variations of the units. It is not feasible to record large and even large databases provide the support for the complexity and combinatory of the language; instead it is need to find a way for optimal coverage of the language[2].

Corpus domain: The domain or focus application of design for CBMTTS is very important since the limited domain can reduce the size of corpus and yet preserve the quality of synthetic speech. Several projects has been developed in restricted domain such as in weather forecasts context[1,3] and talking clock. Since the CBMTTS has not yet being specified for any domain application, so the domain specification will not be taken into the consideration for this project.

Malay speech corpus design: Basically, the concatenation engine used for CBMTTS is an engine that using word-based concatenation engine. This engine first uses word corpus to match the input word as in Fig. 2. If the word is not existed in word corpus, then it will form the word from phoneme corpus. To provide the database support for word concatenation or phoneme concatenation, the speech corpus for word and phoneme has to be designed.

Malay text corpus design: Before the creation of speech corpus, text corpus should be designed first. To design the text corpus, few stages need to take into account[12]:

- Select a source text corpus to fit the target domains
- Analyze the source text corpus to obtain the unit statistics
- Select appropriate prompt subjects from the source text
- Inspect and remove unsuitable sentences

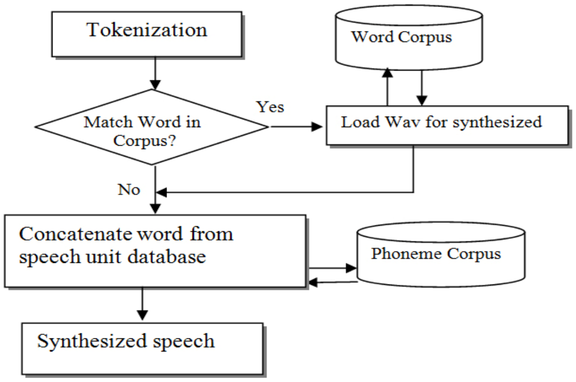


Fig. 2: Process of synthesize speech from word-based concatenate engine

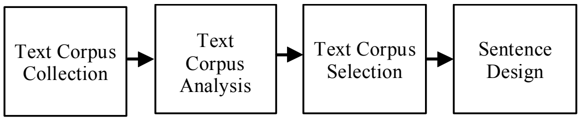


Fig. 3: Text corpus design process

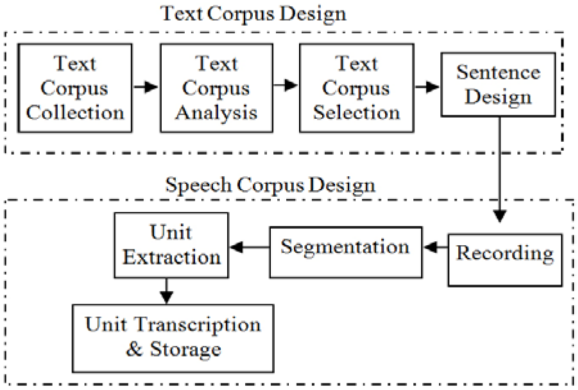


Fig. 4: Corpus building process

Table 1: Detail of 10 million Malay words from internet news

Total text	10027126
Duration of collection	3 months
Total existing Malay word	115738

For Malay text corpus design as shown in Fig. 3, firstly a set of 10 million words of digitized text will be collected from internet news. Then it will go through the process of analysis of the collected text by doing the word frequency count and filter all the non Malay words.

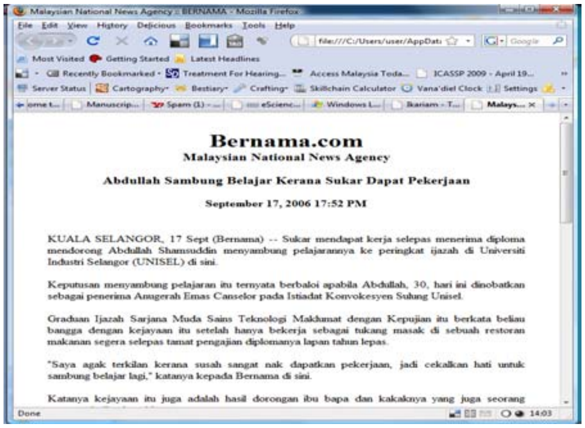
Finally, a set of high frequency word will be used for designing sentences which are for speech corpus design.

Speech corpus design: As shown in Fig. 4, the process of corpus building starts from text corpus design and then speech corpus design.

For text corpus design, a set of online Malay News consisting of 10 million words has been collected in 3 months. Since all the texts are in html format, a text processing tool has been designed to extract text automatically from html file and do the word frequencies count[13]. The information of 10 millions words is shown in Table 1. The design of a well utilized speech corpus requires determining an optimal set of elements for storage in the database[1]. To help in finding an element set of optimal size and composition along these guidelines, some statistical analyses have been conducted[1].

Table 2: 10 millions words resource distribution

Categories	Total	Percentage
Berita Harian	3033569	30.25
Bernama	1218054	12.15
Cybersing	1211249	12.08
Malaysia Kini	2968951	29.61
Utusan Melayu	1595303	15.91
Total sum	10027126	100


Malaysian National News Agency :: BERNAMA - Mozilla Firefox
File Edit View History Delicious Bookmarks Tools Help
file:///C:/Users/user/AppData
Google
Most Visited Getting Started Latest Headlines
Recently Bookmarked Treatment For Hearing... Access Malaysia Toda... ICASSP 2009 - April 19...
Server Status Cartography Bestiary Crafting Skillchain Calculator Vana'diel Clock Settings
ome t... Manuscrip... Spam (1) - ... eScienc... Windows L... Ikariam - T... Malays...
Done
14:03

Bernama.com
Malaysian National News Agency

Abdullah Sambung Belajar Kerana Sukar Dapat Pekerjaan

September 17, 2006 17:52 PM

KUALA SELANGOR, 17 Sept (Bernama) -- Sukar mendapat kerja selepas menerima diploma mendorong Abdullah Shamsuddin menyambung pelajarannya ke peringkat ijazah di Universiti Industri Selangor (UNISEL) di sini.

Keputusan menyambung pelajaran itu ternyata berbaloi apabila Abdullah, 30, hari ini dinobatkan sebagai penerima Anugerah Emas Canselor pada Istiadat Konvokesyen Sulung Unisel.

Graduan Ijazah Sarjana Muda Sains Teknologi Maklumat dengan Kepujian itu berkata beliau bangga dengan kejayaan itu setelah hanya bekerja sebagai tukang masak di sebuah restoran makanan segera selepas tamat pengajian diplomanya lapan tahun lepas.

"Saya agak terkilan kerana susah sangat nak dapatkan pekerjaan, jadi cekalkan hati untuk sambung belajar lagi," katanya kepada Bernama di sini.

Katanya kejayaan itu juga adalah hasil dorongan ibu bapa dan kakaknya yang juga seorang

Fig. 5: html files for the text source from internet news

Text corpus collection and analysis: The 10 millions words from internet news have been collected for few weeks. There are altogether 10,027,126 collected texts as shown in Table 1. The duration for collecting the text is 3 months and altogether 115738 Malay words in the database. The detail of the texts source is listed in Table 2. From Table 2, it can be seen that the text source from Berita Harian is the highest and contributes 30.25 percent. It is followed by Malaysia Kini which contributes 29.61%.

But the text is in ".html" file, so a tool for extracting text from ".html" files has been designed to extract the text and put it into a file and in sentences format as shown in Fig. 5. Then, the sentences are break into words and the words will go through the process of word frequency count.

The result of top 10 highest frequency words is listed in Table 3. It can be seen that the top highest frequency words have been contributing 10.58 percent of the overall 10 millions texts.

Word selection: The CBMTTS utilizes word based concatenation engine if the target word not support in word-corpus, then the word will be constructed through phoneme based unit selection. To provide highest coverage of words for word based concatenation, word selection is the most important issue to be tackled here.

There are many ways of selecting the words for sentences design. One of the way is using words selected from frequency dictionary for example Estonian TTS system using frequency dictionary Estonian, which is based on texts from media and fiction[5]. But there are no available frequency word dictionary for Malay language, thus, the frequency count from 10 million words has been used for this purpose.

The design of text corpus will first focus the word coverage by selecting top highest frequency words that can cover up to 70% of the existing words. To select those words, the word needs to be grouped into few categories such as in Table 4. Then among the groups identify those words that fulfill the requirement of 70% coverage. From Table 4, it can be seen that the word in categories above 1000 have covered 70% of word occurrence. So, those words have been selected to design the carrier sentences. Figure 6 shows the coverage graph for 10 millions words. From this graph, it can be seen that altogether 1451 high frequency words covered 70% of overall words in 10 million collected texts corpus. The word coverage in different word frequency is listed in Table 5. From this table, the word with initial “d” contribute the highest portion of existing word followed by “m” and “s”.

Sentence design: The aim of sentences design is to create a list of text corpus that covers all the selected 1451 words.

Table 3: 10 highest frequency words

	Word	Counting
1	Yang	282200.00
2	Dan	253097.00
3	Untuk	91374.00
4	Tidak	84443.00
5	Pada	60179.00
6	Akan	58222.00
7	Saya	55550.00
8	Kepada	55497.00
9	Mereka	55175.00
10	Ke	42310.00
	Total	1038047.00
	% of 10 millions:	10.58

Table 4: Word coverage according to word frequency groups

Occurrence	Total words	Total occurrence	% coverage
>5000	281	4527844	45.16
2501-5000	319	1115981	11.13
2001-2500	147	327573	3.27
1501-2000	278	480475	4.79
1001-1500	426	521977	5.21
Total	1451	6973850	69.55

able 5: Word frequency count according to word occurrence categories

Word starting with	>5000	2501-5000	2001-2500	1501-2000	1001-1500	Total	% of occurrence
a	314552	63421	24621	23857	24799	451250	6.47
b	344697	107245	31825	47646	46535	577948	8.29
c	5272	15422	10958	5496	6172	43320	0.62
d	793327	53176	13203	27675	30912	918293	13.17
e	12886	3749	4741	5098	8613	35087	0.50
f	7430	0	4023	0	8817	20270	0.29
g	0	10566	4507	6728	11738	33539	0.48
h	58892	32796	13179	8631	10689	124187	1.78
I	299863	19927	8698	8808	9564	346860	4.97
j	92426	44151	4527	9751	10980	161835	2.32
k	368520	141174	17711	51065	59050	637520	9.14
l	134874	26879	4571	14098	18045	198467	2.85
m	437135	157995	37842	74997	82837	790806	11.34
n	61217	12601	4470	10258	5211	93757	1.34
o	59908	0	2137	0	3976	66021	0.95
p	305175	154448	55810	60090	67866	643389	9.23
q	0	0	0	0	0	0	0.00
r	43143	16258	8889	6263	13653	88206	1.26
s	457194	157702	37543	68069	46997	767505	11.01
t	313906	81007	26644	31047	35117	487721	2.12
u	117011	7133	4824	10616	8543	148127	2.12
v	0	0	0	0	0	0	0.00
w	18216	7683	4804	10282	5395	46380	0.67
x	0	0	0	0	0	0	0.00
y	282200	2648	0	0	2393	287241	4.12
z	0	0	2046	0	4075	6121	0.09
Total occurrence Words	4527844	1115987	327573	480475	521977	6973850	100.00
Accumulate occurrence words	452788	56432885	5971398	6451873	6973850	29568790	
Word	281	319	147	278	426	1451	
Accumulate word	281	600	747	1025	1451	1451	

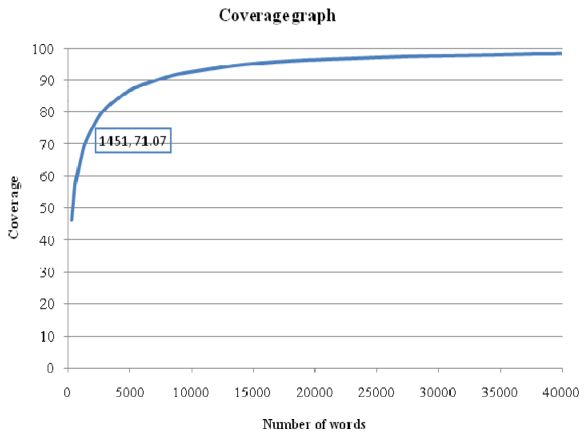


Fig. 6: Coverage graph for 10 million words

RESULTS

Table 6 shows the total units after extract the phoneme units from the carrier sentences. It can be seen that "a" and "e" is two phonemes with highest occurrence in Malay language. They contribute almost 18.36 and 8.6%. The least phoneme is "_z" and "iu". The phoneme speech units has been extracted and grouped in same phoneme folder instead of keep it in original carrier sentence which to reduce the buffer or memory allocation when the system want to access to certain phoneme from its source. The length of sentences may vary from 1.994s to 6.608s and the phoneme is only between 32 and 254ms. To allocate memory for phoneme is smaller than allocate memory for sentences .Then, extract the phoneme from the carrier sentence. For instances, if we intend to form a sentence when consist of 15 phoneme, it may be needed to allocate 15 times of memory for each origin sentences before it can be extracted from the origin source. This will consume a lot of memory and slow down the process of concatenation.

DISCUSSION

The summary of the corpus database for CBMTTS is shown as in Table 7. From this design there are altogether 381 carrier sentences which consist of 16826 phoneme units. All the phone units are phonetically transcript to preserve the phonetic context of its origin that will be used for phonetic context unit selection. The total storage size is 37.6Mb which is bigger than previous Malay TTS diphone database because it uses real wave and support unit selection which provide more choice for each target unit.

Table 6: Total units after extract the phoneme units from the carrier sentences

pho	Total	(%)	pho	Total	(%)
_a	107	0.64	b	253	1.50
_ai	4	0.02	c	77	0.46
_au	1	0.01	d	313	1.86
_b	256	1.52	e	1448	8.60
_c	29	0.17	eh	124	0.74
_d	269	1.60	f	38	0.23
_e	3	0.02	g	169	1.00
_eh	10	0.06	h	374	2.22
_f	17	0.10	i	970	5.76
_g	30	0.18	ia	87	0.52
_h	65	0.39	io	3	0.02
_i	74	0.44	iu	1	0.01
_ia	12	0.07	j	164	0.97
_i	49	0.29	k	665	3.95
_k	248	1.47	kh	7	0.04
_kh	8	0.05	l	514	3.05
_l	72	0.43	m	492	2.92
_m	447	2.66	n	1293	7.68
_n	33	0.20	ng	500	2.97
_ny	2	0.01	ny	91	0.54
_o	21	0.12	o	206	1.22
_p	320	1.90	p	276	1.64
_r	58	0.34	q	1	0.01
_s	258	1.53	r	838	4.98
_sy	5	0.03	s	410	2.44
_t	178	1.06	sy	8	0.05
_u	42	0.25	t	652	3.87
_v	5	0.03	u	696	4.13
_w	18	0.11	ua	107	0.64
_y	59	0.35	ui	2	0.01
_z	1	0.01	v	10	0.06
a	3076	18.3	w	72	0.43
ai	97	0.58	y	52	0.31
au	26	0.15	z	13	0.08

Table 7: Summary of corpus database for CBMTTS

Specification	Detail
Total carrier sentences	381
Total phoneme	16826
Total size	37.6Mb
Storage format	Wave
Sampling frequency	16kHz
Recording environment	Normal room (control noise)
Recording equipment	Multi-speech 4500 (high quality speech DAQ and microphone)
Speaker	Female

CONCLUSION

A set of Malay Speech Corpus has been designed through this project and it is expected to cover around 70% of high frequency words. This corpus may not be the best corpus that supports most of the existing words but it is sufficient to provide the large coverage of words and at the same time keeping the data size at minimum level such as 381 sentences with 37.6 Mb. It can be further enlarged to bigger size and coverage by selecting more words but will be traded off with storage size.

ACKNOWLEDGEMENT

This research project is supported by CBE (Central of Biomedical Engineering) at University Technology Malaysia and funded by Minister of Science and Technology (MOSTI), Malaysia under grant "To Develop a Malay Speech Synthesis System for Standard Platform Compatibility and Speech Compression" Vot 79190.

REFERENCES

1. Nagy, A., P. Pesti, G. Németh and T. Bőhm, 2005. Design issues of a corpus-based speech synthesizer. Hungarian J. Commun., 6: 18-24. www.cc.gatech.edu/~pesti/pubs/ht_cikk_en_2005.pdf
2. Hasim, S., G. Tunga and S. Yasar, 2006. A corpus-based concatenative speech synthesis system for Turkish. Turk. J. Elect. Eng. Comput. Sci., 14: 209-223. http://journals.tubitak.gov.tr/elektrik/issues/elk-06-14-2/elk-14-2-1-0412-1.pdf
3. Fek, M., P. Pesti, G. Nemeth, C. Zainko and G. Olaszy, 2006. Corpus-based unit selection TTS for hungarian. Lecture Notes in Computer Science. In: Text, Speech and Dialogue, 9th International Conferences, Sep. 11-15, Czech Republic, pp: 367-374. http://www.springerlink.com/content/mr6m71133887823m/fulltext.pdf
4. Piits, L., M. Mihkla, T. Nurk and I. Kiissel, 2007. Designing a speech corpus for estonian unit selection synthesis. Proceedings of the 16th Nordic Conference of Computational Linguistics NODALIDA-2007, May 24-26, Tartu, pp: 367-371. www.keeletehnoloogia.ee/projektid/konesyntees/Nodalida2007-syntees.pdf
5. Chou, F.C., C.Y. Tseng and L.S. Lee, 2002. A set of corpus-based text-to-speech synthesis technologies for Mandarin Chinese. IEEE Trans. Speech Audio Process., 10: 481-494. DOI: 10.1109/TSA.2002.803437
6. Rutten, P., G. Coorman, J. Fackrell and B. Van Coile, 2000. Issues in corpus based speech synthesis. IEE Seminar on State of the Art in Speech Synthesis, Apr. 13, Savoy Place, London, pp: 16/1-16/7. http://ieeexplore.ieee.org/xpl/freeabs_all.jsp?arnumber=846973
7. Fung, T.Y. and H.M. Meng, 2000. Concatenating syllables for response generation in spoken language applications. Proceedings 2000 IEEE International Conference on Acoustics, Speech, and Signal Processing. June 5-9. Istanbul, Turkey, pp: II933-II936. DOI: 10.1109/ICASSP.2000.859114
8. Kawai, H. and M. Tsuzaki, 2002. Study on time-dependent voice quality variation in a large-scale single speaker speech corpus used for speech synthesis. Proceeding of the IEEE Workshop on Speech Synthesis. Sep. 11-13, Santa Monica, CA., USA., pp: 15-18. DOI: 10.1109/WSS.2002.1224362
9. Breen and Jackson, 1998. A phonologically motivated method of selecting nonuniform units. International Conference on Speech and Language Processing (ICSLP98), Nov. 30-Dec. 4, Sydney, Australia, http://www.isca-speech.org/archive/icslp_1998/ i98_0389.html
10. Ni, J.F., T. Hirai and H. Kawai, 2006. Constructing a phonetic-rich speech corpus while controlling time-dependent voice quality variability for english speech synthesis. IEEE International Conference on Acoustics, Speech and Signal Processing, May 14-19. Toulouse, France, pp: I-881-I-884. DOI: 10.1109/ICASSP.2006.1660162
11. Gimenez, G.F.M., M.H. Savoji and J.M. Pardo, 1994. New algorithm for spectral smoothing and envelope modification for LP-PSOLA synthesis. Acoustics, Speech, and Signal Processing, IEEE International Conference on, Apr. 19-22, Adelaide, Australia, pp: 573-576. DOI: 10.1109/ICASSP.1994.389229
12. Isogai, M., H. Mizuno and K. Mano, 2005. Recording script design for corpus-based TTS system based on coverage of various phonetic elements. IEEE International Conference on Acoustics, Speech and Signal Processing, Mar. 18-23, Philadelphia, PA., USA., pp: 301-304. DOI: 10.1109/ICASSP.2005.1415110
13. Galicia-Haro, S.N., 2003. Using electronic texts for an annotated corpus building. 4th Mexican International Conference on Computer Science, Sep. 8-12. Tlaxcala, Mexico, pp: 26-32. DOI: 10.1109/ENC.2003.1232870