Research on E-model Enhancement Based on Thai: Enhancement Referring to Packet Loss Effects

Therdpong Daengsi*, Saowanit Sukparungsee, Chai Wutiwiwatchai*** and Apiruck Preechayasomboon******

***Faculty of Information Technology, KMUTNB, Thailand**
****Department of Applied Statistics, Faculty of Applied Science, KMUTNB, Thailand**
*****Human Language Technology Laboratory, NECTEC, Thailand**
******Network Planning Department, TOT, Thailand**

Abstract

This paper starts with a brief look at Thai language and culture, background information about E-model, and a survey on previous works. Then, presents the results from the study of voice quality evaluation referring to packet loss effects using subjective tests with a group of Thai subjects speaking their own language and results from the objective measurement tests called the E-model. The evidence of those results shows that the E-model requires calibration for Thai users. The result from the subjective tests has been used to improve the E-model so that the enhanced E-model can work within Thai environments gaining high accuracy. The outcome of this study is the new factor called the Thai bias factor to improve/enhance the E-model. This approach may be applied to other countries that have their own language and culture.

Keywords : ***VoIP; subjective tests; E-model; packet loss; Thai***

Introduction

Voice over Internet Protocol (VoIP), the modern telecommunication, impacts people around the world because it has several advantages (e.g. extra low rates). However, with different philosophy between data and voice, voice quality problems often arise and become the unwanted consequence of VoIP [1]. To control voice quality of VoIP, many voice quality measurement methods, subjective and objective, have been issued [2]. Practically, it is difficult to conduct the subjective tests. Therefore, objective measurement methods (e.g. E-model), are usually used widely. However, the term 'voice quality' is unclear; different people with different language and culture may give different scores although it is actually the same. Thus, this research, based on representatives of Thai users in Thailand speaking their own language, has been conducted.

Background

Facts about Thai Language and Thai Users

The Thai language is a tonal language, consisting of five tones, used by a population of about 65 million people in Thailand. Different tones in Thai results in different words. Thai is very different from Western languages and other Asian languages (e.g. Chinese, Japanese and Arabic). Tone is about changing the fundamental frequency (F0), therefore, Thai language is an F0 sensitive language [3]. There is evidence that proves Thai subjects have a better perception to Thai sounds than non-Thai native speakers (e.g. English and Chinese) and have also better perception to Thai sounds, than sounds of another language [4-5]. Besides, one of the natures of Thai people is a more compromise cultural structure, comparing to Western people.

E-model

E-model is the most popular objective methods [2, 6]. The output of the E-model is called R-value that can be computed using (1) [7]. R-value can be mapped into the mean opinion score (MOS) [8], using (2). However, it has been stated that *"the E-model has not been fully verified by field surveys or laboratory tests for the very large number of possible combinations of input parameters"* [7].

$$R = Ro - Is - Id - Ie_{eff} + A \tag{1}$$

$$MOS_{CQE} = \begin{cases} 4.5 & ; R>100 \\ 1+0.035R+R(R-60)(100-R)7*10^{-6} & ; 0<R<100 \\ 1 & ; R<0 \end{cases} \tag{2}$$

Where: R = R-value
Ro = signal-to-noise-ratio
Is = simultaneous impairment factor
Id = delay impairment factor
Ie_{eff} = equipment impairment factor, including packet loss
A = advantage factor.
MOS_{CQE} = mean opinion score – estimated conversational quality

Summary

According to cultural and language variation [9-10] and the evidence [4-5], it is supposed that the Thai user perception of voice quality to VoIP is significantly different from the theory that's based on the English language. Moreover, previous works [11-20] have not improved the E-model using subjective results intensively.

Table I. Comparison between MOS_{CQS} and MOS_{CQE}, where Ns is the subject numbers in each test.

Ppl (%)	Ns	MOS_{CQS}	MOS_{CQE}	MOS_{CQS} - MOS_{CQE}
0	26	4.15	4.38	-0.23
4.58	28	3.75	3.02	0.73
9.30	26	3.33	2.15	1.18
19.28	24	2.76	1.45	1.31

Experimental Design and Test Facilities

The design is based on evaluation of voice quality referring to VoIP under random packet loss, both subjective and objective tests. Packet loss rates were 0%, 5%, 10% and 20%, while the codec is G.711 and packet delay was < 10 ms.

For subjective tests, each condition was designed to have 24 subjects who were students in KMUTNB. The subjective method was conversation-opinion tests, using paper based Richard's task [8, 21]. The advantages of conversation-opinion tests are realism and obtaining two sets of data per round. MOS_{CQS} is the final result [22]. For the objective tests, the available E-model tool from the TOT Innovation Institute has been used to obtain the MOS_{CQE} [22]. Finally, the MOS_{CQS} and MOS_{CQE} have been used to find the new factor with Thai users.

The two low background noise rooms in the studio at the Central Library of KMUTNB were used for the conversation task by two participants at the same time. The VoIP system was implemented from Asterisk with two SIP phones. This system run as a direct system that provides the best possible condition, whereas Dummynet was used to generate random packet loss [23].

Results and E-model Enhancement

The subjective results were gained from 49 female and 55 male subjects (15-23 years old). Whereas, the E-model results were from the repetition times of 30 for each condition (the outliers have been found and discarded). The results are in Table I and Figure 1.

For this study, the new factor, called the Thai bias factor, which refers to packet loss effects, will be used to improve the E-model results. However, for E-model, the packet loss impairment factor is referred to Ie_{eff}. Whereas, the language impairment factor has been defined as Il in [14] already. Therefore, the Thai bias factor, including language impairment factor, can be specified as Il_{th}.

Nevertheless, instead of following the way to determine Il as [14], Il_{th} can be found from a simple approach using the differences of MOS_{CQS} and MOS_{CQE}, as in (3):

$$Il_{th} = K(C_1(Ppl)1/3 - C_2)/100 \tag{3}$$

Where: Il_{th} = the Thai bias factor
K = the $F0$ sensitiveness factor that is 5 for Thai because Thai language has 5 tones, while it is 1 for non-tonal languages
Ppl = packet loss probability
C_1, C_2 = the coefficients for Thai language, which are 12.34 and -4.60 respectively.

Eventually, Il_{th} can be combined to the equation (1), then, the enhanced E-model can be presented in (4). However, there are still some errors from the E-model enhancement using Il_{th}, as in Table II.

$$R = Ro - Is - Id - Ie_{eff} + Il_{th} + A \tag{4}$$

Discussion

After calibrating using the MOS_{CQS}, MOS_{CQE^*} has been determined, as in Table II, which shows that MOS_{CQE^*} values become closer to MOS_{CQS} values. However, the equations of E-model and Il_{th} should be improved to obtain more accuracy and reliability because there are still a few errors.

Nevertheless, the available E-model tool that has been used in this study is the only one employed. Therefore, the finding of the second E-model tool to test under the same conditions and facilities is important. From this study, the data collected can be used for research by other countries that have their own cultures and language. Also, the E-model should be calibrated by using the MOS_{CQS} from a group of native speakers to obtain high accuracy and reliability.

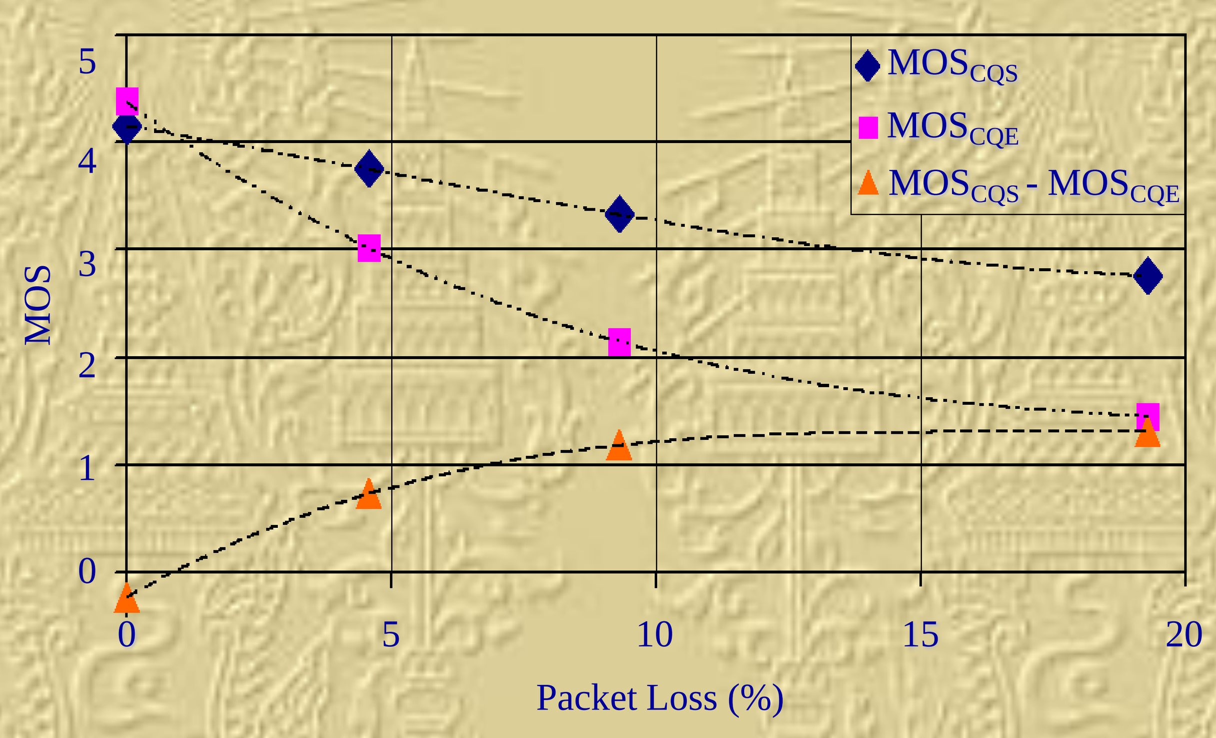

Figure 1. The curves of MOS_{CQS}, MOS_{CQE} and the difference

Conclusion

This research has enhanced the E-model for use in Thai environments particularly, referring to packet loss, using the MOS_{CQS}. Therefore, the enhanced E-model should be able to measure voice quality of VoIP in a Thai environment for Thai users and gain high accuracy. Besides, if this approach is verified and works successfully, it can be applied to other countries that have their own language and culture.

Acknowledgment

This paper is dedicated to HM King Bhumibol Adulyadej of Thailand who is the "Father of Thai Invention," the "Father of Thai Technology," the "Father of Thai Innovation" and the "Father of Royal Rainmaking" on the occasion of his 84 years anniversary.

Gratitude to Dr. Gareth Clayton, my advisor who passed away sadly. Thank you Mr. Gary Sherriff for editing.

References

[1] Avaya Labs., "Avaya IP Voice Quality Network Requirements," Avaya Inc, CO, Apr 2006.
[2] Karapantazis and F.-N. Pavlidou, "Voip: A comprehensive survey on a promising technology," Comput Netw., 2009, vol. 53(12), 2050-2090.
[3] T. Daengsi, C. Wutiwiwatchai, A. Preechayasomboon and S. Sukparungsee, "VoIP Quality Measurement: Insignificant Voice Quality of G.711 and G.729 Codecs in Listening-Opinion Tests by Thai Users," Inform Technol J., In Press.
[4] J. Gandour, D. Wong, L. Hsieh, B. Weinzapfel, D. V. Lancher and G. D. Hutchins, "A crosslinguistic PET study of tone perception," *J Cognitive Neurosci.*, 2000, vol. 12, no. 1, pp. 207-222.
[5] W. Sittiprapaporn, C. Chindaduangratn, and N. Kotchabhakdi, "Brain electric activity during the preattentive perception of speech sounds in tonal languages," *Songklanakarin J. Sci. Techno*l., 2004, vol. 26, no. 4, pp. 439-445.
[6] J. Ren, D. Mao, and Z. Wang, "A neural network based model for VoIP speech quality prediction," in *Proc. Int. Conf. Interaction Sciences*, 2009, pp.1244-1248.
[7] ITU-T Recommendation G.107, "The E-model: a computational model for use in transmission planning," Apr 2009.
[8] ITU-T Recommendation P.800, "Methods for subjective determination of transmission quality", Aug 1996.
[9] M. Al-Akhras, H. Zedan, R. John, and I. Almomani, "Non-intrusive speech quality prediction in VoIP networks using a neural network approach", Neurocomputing, 2009, vol. 72, no. 10-12, pp. 2595-2608.
[10] T. A. Hall, "Objective speech quality measures for Internet telephony," in: Voice over IP (VoIP) Technology, in *Proc. SPIE*, vol. 4522, 2001, pp. 128-136.
[11] A. E. Mahdi and D. Picovici, "Advances in voice quality measurement in modern telecommunications," *Digit Signal Process.*, vol. 19, no. 1, pp. 79–103, 2009.
[12] L. Ding, Z. Lin, A. Radwan, M. El-Hennaway and R. Goubran, "Non-intrusive single-ended speech quality assessment in VoIP," *Speech Commun.*, 2007, vol. 49, pp. 477-489.
[13] A. Meddahi, H. Afifl, and D. Zeghlache, ""Packet-e-model": e-model for wireless VoIP quality evaluation," in *Proc. 14th IEEE Int. Symp. Personal, Indoor and Mobile Radio Communications (PIMRC2003)*, 2003, pp. 2421-2425.
[14] J. Ren, H. Zhang, Y. Zhu and C. Cao, "Assessment of effects of different language in VoIP," in *Proc. Int. Conf. Audio, Language and Image Processing*, 2008, pp. 1624-1628.
[15] J. Ren, C. Zhang, W. Huang and D. Mao, "Enhancement to E-Model on standard deviation of packet delay," in *Proc. Int. Conf. Information Sciences and Interaction Sciences*, 2010, pp. 256-259
[16] M. Voznak, M. Tomes, Z. Vaclavikova and M. Halas, "E-model improvement for speech quality evaluation including codecs tandeming," in *Proc. Int. Conf. DNCOCO '10*, 2010, pp. 119-124.
[17] L. Sun and E.C. Ifeachor, "Prediction of Perceived Conversational Speech Quality and Effects of Playout Buffer Algorithms," in *Proc. IEEE Int. Conf. Comm. 2003 (ICC '03)*, 2003, pp. 1–6.
[18] Y. Bandung, C. Machbub, A. Z.R. Langi and S. H. Supangkat, "Enhancing Low Speecd Network Performance for Voice over Internet Protocol.," in *Proc. Int. Symp. Information Technology*, 2008, pp. 1-6.
[19] A. W. Rix, "Comparison between subjective listening quality and P.862 PESQ score," Psytechnics, Sep 2003.
[20] E.M. Yiu, B. Murdoch, K. Hird, P. Lau and E.M. Ho, "Cultural and language differences in voice quality perception: a preliminary investigation using synthesized signals," Folia Phoniatr Logop, 2008, vol. 60 (3), pp. 107–119.
[21] ITU-T Recommendation P.805, "Subjective evaluation of conversational quality," Apr 2007.
[22] ITU-T Recommendation P.800.1, "Mean Opinion Score (MOS) terminology," Jul 1996.
[23] M. Carbone and L. Rizzo. "Dummynet revisited," *SIGCOMM Comput Commun Rev.*, 2010, vol. 40(2), pp. 12–20.

This research has been sponsored by
Faculty of Information Technology
and Graduate College, KMUTNB;
and Human Language Technology Laboratory, NECTEC.

Table II. Comparison between the MOS_{CQS} and the improved MOS_{CQE} (MOS_{CQE^*}), and Errors of the enhanced E-model

Ppl (%)	MOS_{CQS}	MOS_{CQE^*}	MOS_{CQS} - MOS_{CQE^*} (Error)	Error (%)
0	4.15	4.15	0	0
4.58	3.75	3.78	0.06	1.60
9.30	3.33	3.27	0.11	3.30
19.28	2.76	2.84	0.11	3.99