

E-Model Modification for Multi-Languages over IP

Tuul Triyason¹, Prasert Kanthamanon¹

¹*School of Information Technology, King Mongkut's University of Technology Thonburi, 126, Pracha Uthit Rd., Bang Mod, Thung Khru, Bangkok, Thailand 10140
tuul.tri@sit.kmutt.ac.th*

Abstract—It is well known that non-intrusive speech quality assessment methods are appropriate for real time monitoring of VoIP traffic. However, previous researches has proved that most of the non-intrusive speech quality assessment methods failed to estimate accurate speech quality using different languages. Consequently, intrusive methods are frequently chosen to provide a more accurate measurement, however they cannot be used for real time VoIP traffic monitoring. In this paper, the technique to enhanced simplified version of ITU-T recommendation G.107 E-model with a language impairment parameter was proposed. The method to estimate the function of language impairment by tuning the E-model with an intrusive objective method, PESQ was presented. The results from statistical analysis show that the modified E-model matches well with PESQ scores in eight languages.

Index Terms—VoIP, speech quality prediction, E-model, PESQ.

I. INTRODUCTION

The penetration of high speed mobile Internet access has created the current trend of speech communication over IP networks. Voice over Internet Protocol (VoIP) is a technology that carries voice data in the form of a packet and sends it across the IP network. VoIP has become extremely popular because it is free or low cost when compared to traditional telephone services. However, VoIP is a delay sensitive application that relies on the best effort network. The speech quality can be impaired by the characteristics of the communication channel (e.g., delay, variation of delay and packet loss). In order to maintain a good Quality-of-Service (QoS) to meet a commercial agreement, an efficient accurate speech quality assessment method is required.

In the past, speech quality assessment methods have been developed with either subjective or objective techniques. Subjective measurement is an assessment made by human being and the quality is rated by Mean Opinion Score (MOS) as presented in ITU-T Recommendation P.800 [1]. The method asks the subjects to rate the quality by listening to a set of speech samples under a controlled environment. The quality scores will range from 1 (Poor) to 5 (Excellent) scales. The subjective test can be conducted by a listening or conversational test. If the test is a listening test, MOS_{LQ} is

used as the listening quality index. If the test is a conversational test, MOS_{CO} is used as the conversational quality index. It is well known that subjective testing is the most accurate and reliable method for assessing speech quality. However, this method still has some limitations. First, it is a time consuming and also an expensive method. Second, the results from the same subject may change if he/she is made to repeat the experiment. These drawbacks of subjective testing have led to the development of objective measurement methods.

Objective measurement can be either an intrusive or a non-intrusive method. The intrusive method typically compares two input signals, a reference signal and a degraded signal which are received at the receiving side of the communication chain. The most successful intrusive method of ITU-T is defined in recommendation P.862 (PESQ) in 2001 which was used for narrow-band measurement [2]. Later, it was complemented with P.862.2 for wideband measurement [3]. Compared to non-intrusive methods, most of the intrusive methods have more accuracy but are inconvenient to use for real time monitoring. The non-intrusive methods focus on measuring real-time voice traffic. This type of measurement is designed to predict speech quality without a reference signal. The most popular non-intrusive method of ITU-T is defined in recommendation G.107 E-model which was designed as a transmission planning tool [4]. The E-model is a mathematical model that merges the impairment factors on the conversational path, before calculate conversational quality.

However, the previous researches has proved that non-network impairments such as language can slightly impact on the accuracy of objective methods [5]–[8]. Conversely, most of objective measurement studied only deals with the impact of network impairment parameters such as codec, packet loss, delay and jitter [9]–[12]. To the best of our knowledge, there are a few works that pay attention to improve the objective method E-model with non-network parameter. The objective of this work is to provide the method to enhance the simplified version of the E-model with a language impairment parameter. The remainder of the paper is organized as follows: In Section II, the details of the simplified E-model are presented. In Section III, the effect of language on objective measurement and the concept of language impairment parameter are presented. In Section IV,

Manuscript received July 16, 2014; accepted October 24, 2014.

This research was supported by the Higher Education Research Promotion and National Research University Project of Thailand, Office of the Higher Education Commission.

the statistical evaluation of the enhanced simplified E-model and results are reported. The conclusion is presented in Section VI.

II. THE SIMPLIFIED E-MODEL

The original E-model, defined in ITU-T recommendation G.107 is a mathematical model used for network planning purposes [4]. The E-model combines network impairment factors from circuit switch network and packet switch network before converting to measure speech quality R factor, ranging from 0 (poor) to 100 (excellent). The impairment parameters included in the original E-model are merged among network impairment factors such as delay, signal loss, background noise, echo and others. This made the original E-model equation too complicated [13]. The original E-model from ITU-T proposed shown in (1)

$$R_{\text{conversation}} = R_0 - I_s - I_d - I_{e\text{-eff}} + A, \quad (1)$$

where R_0 represents the basic signal to noise ratio at 0 dB_r, I_s is a sum of all impairments which may occur in voice transmission. I_d is the impairment factor representing all impairments due to delay. $I_{e\text{-eff}}$ or effective impairment equipment parameter is a combination between impairment equipment parameter at zero packet loss (I_e) and function of I_e that is dependent on packet loss level. A is an expectation factor which explains lower customer expectations of quality. Most parameters in the original E-model are specific values which are available in the ITU-T recommendation G.113 [14]. According to AT&T lab, some parameters of the E-model are not related to a packet switch network [15]. Therefore, the simplified version of the E-model was proposed as shown in (2)

$$R_{\text{conversation}} = R_0 - I_d - I_{e\text{-eff}}. \quad (2)$$

This simplified version focuses on measuring conversational quality only from a packet network. The more details pertaining to this version of the E-model can be found in the AT&T lab paper. Moreover, there are more simplified versions of the E-model as shown in [13], [16], however they are beyond the scope of this work. The detail of each parameter in the simplified E-model are described as follows:

A. Basic Signal-to-Noise Ratio, R_0

R_0 represents the basic signal to noise ratio at 0 dB_r point. According to ITU-T Rec. G.107, if all calculation parameters are set to default values, the R_0 is equal to 93.2 [4]. In this paper, the value of R_0 is set to the same value as in Rec. G.107.

B. Equipment Impairment Factors, I_e

The equipment impairment factor is the impairment caused by low bit-rate coding. The values of I_e from ITU-T codec were presented in ITU-T Rec. G.113 Appendix I [14]. If codec operate under a network with independent or burst packet loss, the I_e value can be described in the relation of effective equipment impairment factor $I_{e\text{-eff}}$. ITU-T Rec. G.107 defined as an equation of $I_{e\text{-eff}}$ as shown in (3)

$$I_{e\text{-eff}} = I_e + (95 - I_e) \cdot \frac{Ppl}{\frac{Ppl}{BurstR} + Bpl}, \quad (3)$$

where the factor I_e is the equipment impairment factor at zero packet loss condition, PPL is the packet loss probability. $BurstR$ is the ratio between average length of loss in observed sequence of packet and average length of bursts expected in the network under random loss condition. In our experiment scenario, $BurstR$ is equal to 1. Bpl is the packet loss robustness factor which also has a specific value for each codec. The packet loss robustness factor Bpl is also listed in Appendix I of ITU-T Rec. G.113 [14].

C. Delay Impairment Factor, I_d

Delay impairment factor I_d is the impairment factor illustrates all impairments from the delay of communication chain. I_d can be calculated through a complicated set of equations as presented in Rec. G.107. For the simplified E-model of AT&T lab, the value of I_d can be obtained from one-way delay in communication path. The relation between one-way delay and I_d is shown in (4) as follows

$$I_d = 0.024d + 0.11(d - 177.3)H(d - 177.3), \quad (4)$$

where $H(x) = \begin{cases} 0, & \text{if } x < 0 \\ 1, & \text{if } x \geq 0 \end{cases}$. Equation (4) shows that in the case of a network with small one-way delay (d less than 100 ms), the effect of delay impairment is not significant and can be discarded [17]. The simplified E-model can be reduced to a more simplified form as shown in (5)

$$R_{\text{conversation}} \approx R_{\text{listening}} = R_0 - I_{e\text{-eff}}. \quad (5)$$

Equation (5) represents the case of small one-way delay, the perceived quality of listening and conversation can be estimated to be the same value. This estimation was supported by the experiment in [18]. The experiment shows that, for VoIP with only packet loss impairment, the subjective conversational test can be replaced by a listening test.

D. R-factor and Transform Function

The R factor can be further transformed into MOS scale with a transform function. ITU-T Rec. G.107 defines the transform function between R and MOS as shown in (6):

$$\begin{cases} (R \leq 6.5) : MOS = 1, \\ (6.5 \leq R \leq 100) : MOS = \\ = 1 + 0.035R + R(R - 60)(100 - R) \cdot (7 \times 10^{-6}), \\ (R > 100) : MOS = 4.5. \end{cases} \quad (6)$$

However, (6) cannot apply to obtain R values from MOS. The transform from R to MOS value can be done by using the Candono's Formula as shown in [19]. In this paper, the transform function of MOS to R value can be obtained by inverting graph of (6) and fitting the curve of a 3rd-order polynomial equation. The polynomial transform function of MOS to R is shown in (7)

$$R = 3.113MOS^3 - 26.105MOS^2 + 89.31MOS - 59.293. \quad (7)$$

III. LANGUAGE IMPAIRMENT PARAMETER

In this section, the procedure for developing a language impairment factor is described. The method of embodiment is shown step-by-step as follows:

A. Experimental Testbed

To consider the effect of language on objective speech quality measurement, an experimental testbed was setup with intrusive method (PESQ) and non-intrusive method (E-model). The testbed was designed as shown in Fig. 1.

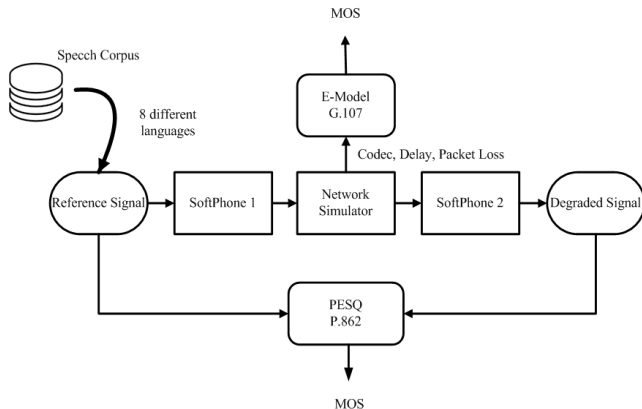


Fig. 1. Experimental testbed design diagram.

TABLE I. EXPERIMENTAL TRAINING CONDITION.

Languages	Packet Loss	Delay	Codec
Arabic, Chinese, Dutch, English, French, German, Japanese, and Thai	0 %–30 % independent loss	<100 ms	G.729

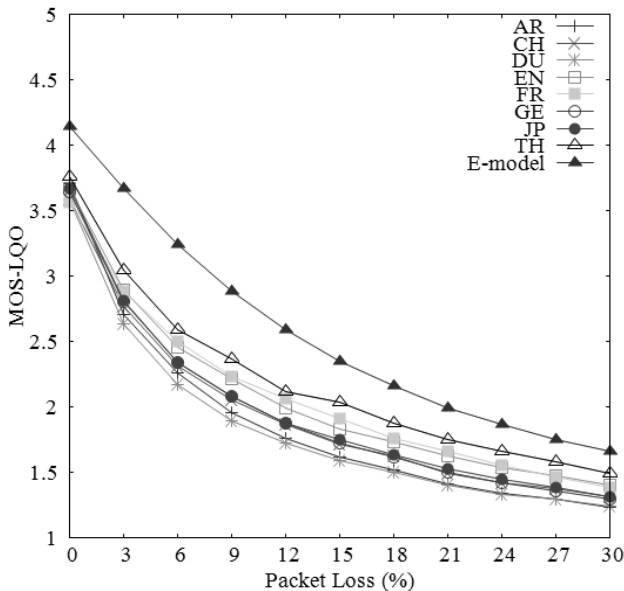


Fig. 2. Effect of Language on PESQ and E-model MOS with G.729 codec under 0 %–30 % packet loss rate.

Two data sets of speech corpus were used as reference speech files in this experiment. The first reference set came from ITU-T recommendation P.50 speech corpus [20]. It contains seven languages (British English, Arabic, Chinese, Dutch, French, German and Japanese), 8 second–12 second duration speech files, which represent speech from sixteen

talkers, eight male and eight female. The later reference set comes from the *Thai Speech Set for Telephony (TSST)* corpus [21]. It contains Thai language from four adult females, four adult males, and two children. However, the children speech files were excluded from the experiment as mentioned in [22]. All speech files were prepared sampling rate to 8 kHz, 16 bit linear PCM. The voice packet was created by softphone1 and softphone2, which was developed from the PJSIP stack library [23]. The speech files were coded with a G.729 codec with default packet size. In order to measure the variation of MOS, 0 %–30 % independent packet loss (3 % increment) was generated from the Network Simulator [24]. The average MOS was calculated from PESQ and E-model. The average one-way delay of the testing environment was lower than 100 ms, hence the delay impairment factor of E-model can be neglected. To avoid the effect from loss position and gender as mentioned in [8], MOS was obtained from the average of 10 measurements of both genders. The detail of training condition and result are shown in Table I and Fig. 2.

B. Language Impairment Parameter

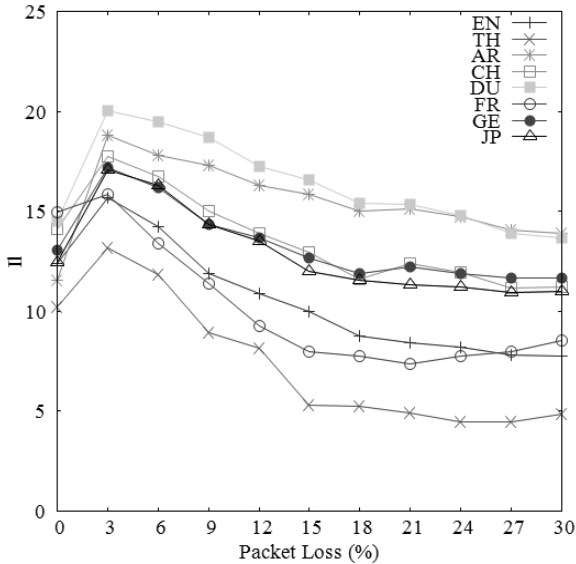
The concept of the language impairment parameter was first introduced to the E-model in [25]. However, the method to define the language impairment function and the coefficient is still not clear. The concept is based on the assumption that the effect from a language should be applied to speech quality as the impairment from the new codec. With this assumption, the E-model with language impairment parameter can be rewritten as shown in (8)

$$R = R_0 - I_{e-eff} - I_l, \quad (8)$$

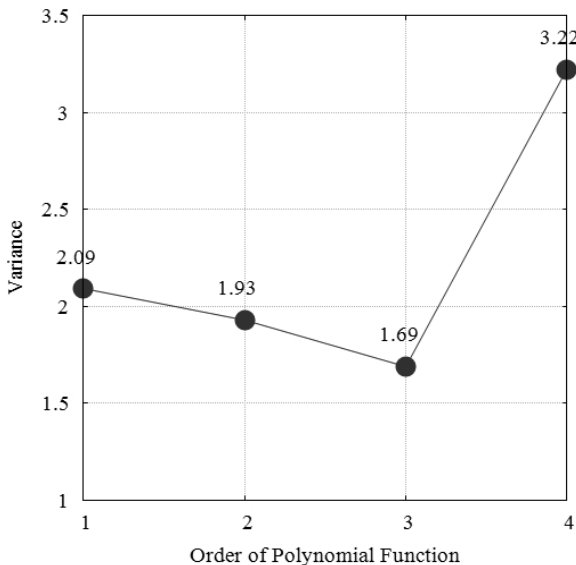
where I_l is the language impairment parameter. By converting PESQ MOS to R factor as described in (7), I_l can be written in the form of the different between effective equipment impairment factors of PESQ and Rec. G.107 E-model as shown in (9)

$$I_l = (I_{e-eff})_{PESQ} - (I_{e-eff})_{G.107}, \quad (9)$$

Now, I_l can be expressed as the difference between I_{e-eff} that has been observed from PESQ and G.107. The impact of this difference is also varied as a function of packet loss increase. In the case of network without packet loss, impairment I_l will be equal to some constant values for each language. When considering the network with packet loss, impairment I_l will be defined by some functions that are dependent on packet loss level. To derive the function of I_l , the PESQ results from Fig. 2 were converted into I_{e-eff} and substituted to (9). The result is the language impairment of each language, which can be plotted in the form of function with packet loss level as shown in Fig. (3). It can be seen that I_l has a greater value when the packet loss rate is about 3 % and much lower when the packet loss is greater than 10 %. The result can imply that, on the network with low level of packet loss, the impact from language impairment parameters is the main factor of difference between I_{e-eff} from the PESQ and the standard E-model.

Fig. 3. I_1 values of eight languages in G.729 codec.

The function of I_l can be described by a polynomial equation with the least square method. The higher order of polynomial function may result in the higher goodness of fit (R^2) value but the curve may result in overfitting behaviour. Conversely, the lower order of polynomial may result in low complexity of equation but the predicted function may result in lower accuracy also. To avoid the overfitting behaviour, an optimum polynomial order was chosen by the variance analysis technique [26].

Fig. 4. Variance analysis of 1st to 4th order polynomial function of I_l .TABLE II. PARAMETERS FOR 3RD REGRESSION MODEL OF EIGHT LANGUAGES IN G.729 CODEC.

Languages	A	b	c	d
Arabic	0.0017	0.0872	1.1432	13.309
Chinese	0.0011	0.0484	0.3599	15.354
Dutch	0.0016	0.0797	0.9279	15.896
English	0.0011	0.0459	0.2221	13.608
French	0.0007	0.0135	0.4297	15.911
German	0.0011	0.0516	0.4442	14.382
Japanese	0.0014	0.0658	0.6009	13.867
Thai	0.0013	0.0872	1.1432	13.309

The order of polynomial with a minimum or no significant

decrease should be chosen. The method of variance calculation is shown in (10) as follows

$$\text{Var}(m) = \frac{\text{SSR}(m)}{n - m - 1}, \quad (10)$$

where $\text{Var}(m)$ is the variance of polynomial order m , $\text{SSR}(m)$ is sum square of residuals from polynomial order m , and n is the number of data point. The I_l function of eight languages were analysed with the variance value of 1st to 6th order polynomial equation. The average variance value of eight languages shows the minimum value at 3rd order polynomial equation with the goodness of fit (R^2) equal to 0.90. The result of variance analysis from 1st to 4th order polynomial equation is shown in Fig. 4 (for 5th and 6th order, the variance values increased dramatically and are not shown in the figure). According to the variance result, the function of I_l was described by the 3rd order polynomial function. Equation (11) shows the general form of the 3rd order polynomial function

$$I_l = aP^3 + bP^2 + cP + d, \quad (11)$$

where, P is percent of random packet loss in network. The constant parameters of each language are described as shown in Table I.

IV. EVALUATION AND RESULTS

In this section, the performance of E-model with language impairment parameter was evaluated. The testing environment was setup in the same scenario as in training condition; however, the packet loss conditions were changed to 0 %, 3 %, 5 %, 8 %, 10 %, 13 %, 15 %, 18 % and 20 %. The concluding conditions of the testing environment are described in Table III.

TABLE III. MODEL TESTING CONDITION.

Languages	Packet Loss	Delay	Codec
Arabic, Chinese, Dutch, English, French, German, Japanese, and Thai	0 %, 3 %, 5 %, 8 %, 10 %, 13 %, 15 %, 18 % and 20 % independent loss	<100 ms	G.729

To avoid the bias from the speaker and the wording of sentences in the speech sample, the different speech corpora were used as a new source of speech file samples. The Chinese, Dutch, British English, French, German, and Japanese samples were selected from the ITU-T P.501 speech repository. The Arabic samples were selected from BBN/AUB Arabic speech corpus of LDC [27]. The Thai language was selected from Lotus speech corpus [28]. Each language contains four sentences from four speakers (two males and two females). All speech sample files were resampling to 8 kHz, 16 bits linear PCM. MOS of eight languages were measured by PESQ, G.107 E-model, and EL-model (E-model with language impairment enhanced). The accuracy evaluation method can be described into two steps. The first step is to check the correlation between MOS of the two E-models with PESQ. In this study, the Pearson correlation coefficient was used as the correlation measurement technique. It is the most widely used for

measuring the correlation between two variables. The value closer to 1 is the more positive correlation (-1 is a negative correlation and 0 is no correlation). The results of the correlation measurement are shown in Table IV.

TABLE IV. CORRELATION COEFFICIENT.

Languages	G.107 E-model	EL-Model
Arabic	0.9730	0.9987
Chinese	0.9704	0.9937
Dutch	0.9724	0.9966
English	0.9848	0.9987
French	0.9902	0.9979
German	0.9646	0.9904
Japanese	0.9630	0.9915
Thai	0.9606	0.9959

From Table IV, the average correlation coefficient of the G.107 E-model of eight languages is 0.9724, whereas the average from the EL-model is 0.9954. It can be observed clearly from the correlation coefficient that the MOS of the EL-model have a more positive relation with PESQ (about 2.3 % improvements from G.107 E-model). However, the correlation coefficient does not take the accuracy of individual components into account. To evaluate the error of prediction, further step of accuracy evaluation was conducted. The error of prediction can be analysed by residual analysis between the predicted values and the actual values. Mean Absolute Percent Error (MAPE) and Root Mean Square Error (RMSE) are normally used to measure the error in this task [29]. MAPE measures the average magnitude of error in the prediction set. It is the average over the verification sample of the absolute values of the differences between forecast and actual. Conversely, RMSE used a quadratic scoring rule which measures the average magnitude of the error. The difference between forecast and corresponding actual value in RMSE are each squared and then averaged over the sample before the square root of the average is taken. The method of MAPE and RMSE calculation can be derived as shown in (11) and (12):

$$MAPE = \frac{100\%}{n} \sum_{t=1}^n \left| \frac{A_t - F_t}{A_t} \right|, \quad (12)$$

$$RMSE = \sqrt{\frac{\sum_{t=1}^n (A_t - F_t)^2}{n}}, \quad (13)$$

where A_t is an actual value, F_t is a forecast value, and n is the number of the sample. The errors of prediction in eight languages were calculated by MAPE and RMSE methods. The results of error analysis are shown in Table V and Table VI.

TABLE V. MAPE ANALYSIS RESULTS.

Languages	G.107 E-model	EL-Model
Arabic	29.7193	3.1224
Chinese	31.5374	2.8280
Dutch	29.7179	9.6273
English	17.9937	6.0032
French	17.2650	4.9864
German	26.5042	4.7091
Japanese	36.1661	4.0193
Thai	39.0836	4.1781

MAPE and RMSE analysis shows that the EL-model has a less error than G.107 E-model in eight languages. The average error reduction of the EL-model is about 80 % improvement (81 % accuracy improvement in RMSE and 78 % in MAPE). This implies that the modified E-model has a higher level of performance in predicting the speech quality with the effect of language impairment in the same level as PESQ.

TABLE VI. RMSE ANALYSIS RESULTS.

Languages	G.107 E-model	EL-Model
Arabic	0.6644	0.1021
Chinese	0.6878	0.0818
Dutch	0.6615	0.2041
English	0.4463	0.1513
French	0.4398	0.1372
German	0.6085	0.1336
Japanese	0.7618	0.1117
Thai	0.7956	0.1033

V. CONCLUSIONS

In this paper, the methodology to enhance the simplified E-model with a language impairment factor was proposed. Based on the concept of language impairment factor, the 3rd order polynomial regression model and coefficients of eight languages in G.729 codec are proposed for I_1 prediction. The modified E-model and the original G.107 E-model were evaluated the accuracy with the PESQ MOS. The new modified E-model can provide an accurate prediction of impairment from language as PESQ nonintrusively. It can be used as a real time monitoring task instead of an intrusive method. The first author would like to point out the two main contributions of this paper. First, the experiment has confirmed that the E-model still has some limitations with impairment from language and its result does not correspond with PESQ. Second, the method to synthesize the new impairment functions and enhance the original E-model. This method can be extended to include other non-network impairment parameters (gender and age) or apply to other codec also.

ACKNOWLEDGMENT

The first author would like to gratefully and sincerely thank M.Sc. IT 27&28 students from KMUTT for their encouragement throughout this work. It is also my duty to record my thankfulness to Dr. Therdpong Daengsi, Dr. Pongpisit Wuttidittachotti, Ms. Nalakhana Khitmoh and Mr. John Lawry who always helped with the data evaluation technique and editing.

REFERENCES

- [1] *Methods for Subjective Determination of Transmission Quality*, ITU Rec. P.800, Int. Telecommun. Union, 1996.
- [2] *Perceptual Evaluation of Speech Quality (PESQ), An Objective Method for End-to-end Speech Quality Assessment of Narrow-band Telephone Networks and Speech Codecs*, ITU Rec. P.862, Int. Telecommun. Union, 2001.
- [3] *Wideband Extension to Recommendation P.862 for the Assessment of Wideband Telephone Networks and Speech Codecs*, ITU Rec. P.862.2, Int. Telecommun. Union, 2007.
- [4] *The E-model: a computational model for use in transmission planning*, ITU Rec. G.107, Int. Telecommun. Union, 2007.
- [5] Z. Cai, N. Kitawaki, T. Yamada, S. Makino, "Comparison of MOS

- evaluation characteristics for Chinese, Japanese, and English in IP telephony”, in *4th Int. Universal Communication Symposium (IUCS 2010)*, 2010, pp. 112–115. [Online]. Available: <http://dx.doi.org/10.1109/IUCS.2010.5666762>
- [6] T. Triyason, P. Kanthamanon, “Perceptual evaluation of speech quality measurement on speex codec VoIP with tonal language Thai”, in *Advances in Information Technology*, Springer Berlin Heidelberg, 2012, pp. 181–190. [Online]. Available: http://dx.doi.org/10.1007/978-3-642-35076-4_17
- [7] D. U. Ebem, J. G. Beerends, J. Van Vugt, C. Schmidmer, R. E. Kooij, J. O. Uguru, “The impact of tone language and non-native language listening on measuring speech quality”, *JAES*, vol. 59, no. 9, pp. 647–655, 2011.
- [8] L. Sun, “Speech quality prediction for voice over Internet protocol networks”, Ph.D. dissertation, School of Computing, Communications and Electronics, Faculty of Technology, Plymouth Univ., UK, 2004.
- [9] Z. Bojovic, Z. Peric, V. Delic, E. Secerov, M. Secujski, V. Senk, “Comparative analysis of the performance of different codecs in a live VoIP network using SIP protocol”, *Elektronika ir Elektrotehnika*, no. 1, pp. 37–42, 2012.
- [10] L. Ding, R. A. Goubran, “Assessment of effects of packet loss on speech quality in VoIP”, in *Proc. 2nd IEEE Int. Workshop on Haptic, Audio and Visual Environments and Their Applications, (HAVE 2003)*, 2003, pp. 49–54.
- [11] J. Holub, O. Slavata, “Impact of IP channel parameters on the final quality of the transferred voice”, in *Wireless Telecommunications Symposium (WTS 2012)*, 2012, pp. 1–5. [Online]. Available: <http://dx.doi.org/10.1109/WTS.2012.6266108>
- [12] Z. Qiao, R. K. Venkatasubramanian, L. Sun, E. C. Ifeachor, “A new buffer algorithm for speech quality improvement in VoIP systems”, *Wirel. Pers. Commun.*, vol. 45, no. 2, pp. 189–207, 2008. [Online]. Available: <http://dx.doi.org/10.1007/s11277-007-9408-7>
- [13] C. Jiang, P. Huang, “Research of monitoring VoIP voice QoS”, in *Int. Conf. Internet Computing Information Services (ICICIS)*, 2011, pp. 499–502.
- [14] *Transmission Impairments due to Speech Processing*, ITU Rec. G.113, Int. Telecommun. Union, 2007.
- [15] R. G. Cole, J. H. Rosenbluth, “Voice over IP Performance Monitoring”, *SIGCOMM Comput. Commun. Rev.*, vol. 31, no. 2, pp. 9–24, 2001. [Online]. Available: <http://dx.doi.org/10.1145/505666.505669>
- [16] H. Assem, D. Malone, J. Dunne, P. O’Sullivan, “Monitoring VoIP call quality using improved simplified E-model”, in *2013 Int. Conf. Computing, Networking and Communications (ICNC)*, 2013, pp. 927–931.
- [17] M. Voznak, “E-model modification for case of cascade codecs arrangement”, *Int. Journal of Mathematical Models and Methods in Applied Science*, vol. 5, no. 8, pp. 1439–1477, 2011.
- [18] L. Gros, N. Chateau, “An impact of listening and conversational situations on speech perceived quality for time-varying impairments”, in *Proc. 1st Int. Conf. Measurement of Speech and Audio Quality in Networks (MESAQIN 2002)*, 2002.
- [19] C. Hoene, H. Karl, A. Wolisz, “A perceptual quality model intended for adaptive VoIP applications: research articles”, *Int. J. Commun. Syst.*, vol. 19, no. 3, pp. 299–316, 2006. [Online]. Available: <http://dx.doi.org/10.1002/dac.755>
- [20] *Artificial Voice*, ITU Rec. P.50, Int. Telecommun. Union, Sep. 1999.
- [21] T Daengsi, A Preechayasomboon, S Sukparungsee, C Wutiwiwatchai, “Thai text resource: a recommended Thai text set for voice quality measurements and its comparative study”, *KKU Science Journal*, vol. 40, no. 4, 2012, pp. 1114–1127.
- [22] K. Yochanang, T. Daengsi, T. Triyason, P. Wuttidittachotti, “A comparative study of VoIP quality measurement from G.711 and G.729 using PESQ and Thai speech”, in *Advances in Information Technology*, pp. 242–255, 2013.
- [23] PJSIP – Open Source SIP Media, and NAT traversal library. [Online]. Available: <http://www.pjsip.org>
- [24] IPNetSim – (IP Network/Wan Emulator). [Online]. Available: <http://www.gl.com/ipnetsim.html>
- [25] J. Ren, H. Zhang, Y. Zhu, C. Gao, “Assessment of effects of different language in VOIP”, in *Int. Conf. Audio, Language and Image Processing, (ICALIP 2008)*, 2008, pp. 1624–1628.
- [26] A. Kaw, E. Kalu, Numerical Methods with Application: Customized for University of South Carolina, 2012. [Online]. Available: Autarkaw.com
- [27] Linguistic Data Consortium (LDC) catalog number LDC2005S08. [Online]. Available: <http://www.ldc.upenn.edu/2005>
- [28] C. Patcharikra, S. Treepop, K. Sawit, T. Nattanun, W. Chai, “LOTUS: Large vocabulary Thai continuous Speech Recognition Corpus”, in *NSTDA Annual Conf. S&T in Thailand: Towards the Molecular Economy*, 2005.
- [29] R. Nicholas, “Predicting the future: An introduction to the theory of forecasting”, State University of New York Press, 1998.
- [30] L. Sun, E. C. Ifeachor, “Voice quality prediction models and their application in VoIP networks”, *IEEE Trans. Multimedia*, vol. 8, no. 4, pp. 809–820, 2006. [Online]. Available: <http://dx.doi.org/10.1109/TMM.2006.876279>
- [31] *Test signals for use in telephony*, ITU Rec. P.501, Int. Telecommun. Union, 2009.