ABSTRACT

Voice service being the major offering of telecommunication networks, its level of Quality of Service (QoS) largely determines the performance of these networks. This work evaluated the state-of-the-art Perceptual Evaluation of Speech Quality (PESQ) objective model for perceptual estimation of the quality of transmitted speech signals. Perceptual estimation of the quality of speech is predominantly done by subjective techniques and the results presented as Mean Opinion Scores (MOS), which has a scale from 1 for poor quality to 5 for excellent quality. Despite constraints of the subjective approach to perceptual speech quality estimation, its scores serves as the basis for correlating quality scores from objective techniques for speech quality estimation. Original or reference speeches were recorded using professional studio equipment and software, and guided by provisions of ITU-T P.830. The speeches were transmitted over three mobile wireless networks. A speech database consisting of 64 original (32 male and 32 female) and 192 transmitted speeches was developed. Reference speeches and their corresponding transmitted (network-degraded) speeches were tested on the PESQ model to estimate their quality scores. The raw PESQ quality scores are within the scale range of -0.5 and 4.5. They were mapped to the
MOS scale for linear comparison of the scales. Study of PESQ model showed several shortcomings, some of which have been improved upon by previous researchers. Evaluating PESQ mapping function (in ITU-T Rec P.862.1) showed the need for better coverage of the MOS scale. Analysis of solution for the logistic growth function was done and parameters were optimised which resulted in the development of a new robust logistic mapping function. The raw PESQ quality scores were mapped using the developed mapping function as well as two known standard mapping functions, namely: ITU-T P.862.1 and Morfitt and Cotanis mapping functions. The mapped scores known as PESQ MOS-listening quality objective (PESQ MOS-LQO) obtained with the three functions were tested using ANOVA at a significant figure of $\alpha = 0.05$. The developed logistic mapping function offered a quality score coverage of 98.6% of the MOS scale. This was evaluated against the two known standard mapping functions and the developed function offered improvement of 11.8 and 4.9% over and above their 86.8 and 93.7% coverage of the MOS scale respectively. At the significance level of $\alpha = 0.05$, an F-value of 60.6042, a critical-F of 3.04, and a p-value of $4.61721E-21$ were obtained. With $p < 0.05$, the Null Hypothesis was rejected, and the critical-F value being less than the F-statistic value confirmed the rejection. Therefore, the data distribution of at least one of the functions has a different mean and belongs to a separate population of performance.

Keywords: Mapping; speech quality; logistic functions; perceptual models; sigmoid symmetry.

1. INTRODUCTION

Assessing the quality of processed or transmitted speech signals in comparison with the reference or original speech from perceptual perspectives or users’ viewpoint has its root in subjective speech quality evaluation techniques standardised as ITU-T Rec. P.830 [1]. Though the subjective measure is bedeviled by several constraints of high cost, being very slow, results being highly variable and not easily reproducible, and so on [2,3,4,5,6,7], yet it is the basis for correlating objective quality estimation measures of all types. This is because well–controlled subjective quality test is been adjudged the most accurate and reliable means of assessing speech quality [8]. Quality of service (QoS) is actually more about the user of a service and his/her satisfaction with the rendering of the service by the provider. This necessitates the definition of service quality as: “The collective effect of service performance that determines the degree of satisfaction of a user of the service.” [9].

In subjective quality tests, listeners (subjects) are engaged to listen to system- or network-degraded speeches and rate the speech quality based on their perception. These quality scores are aggregated and the mean obtained and denoted as the MOS value. Subjective quality testing is carried out using the listening-only technique based on the Absolute Category Rating (ACR) scale ranging between 1 for bad quality and 5 for excellent quality. This subjective MOS quality rating is a well-established scale which [4,10] noted has been applied to both analogue and digital telephone connections and devices such as codecs, for characterising the quality of telephony equipment and services.

Perceptual Evaluation of Speech Quality (PESQ) model like any other objective quality measuring technique was designed with an output score that is different from the subjective MOS score. The raw quality score of PESQ is within the range of -0.5 to 4.5 [11]. The raw PESQ score is mapped to PESQ MOS-LQO, which is relative to the scale of the Subjective MOS. The mapping is done using the mapping function found in the amendment to PESQ standardised as ITU-T Rec. P.862.1 [12]. After mapping, the PESQ MOS-LQO is then correlated with the subjective MOS score to enable us to determine the figure of merit of such objective quality estimation.

2. REVIEW OF PESQ’s LIMITATIONS

PESQ algorithm is robust and has been widely used for assessment of the quality of processed and transmitted speech signals, and for optimisation of telecommunication networks. Nevertheless, a number of research efforts have critically evaluated the algorithm to discover limitations and constraints that made it not to be too accurate under certain conditions. Such efforts have led to modifications and improvements of some aspects of the algorithm. These include efforts at correcting time and level alignment problems, signal spectrum mismatch, mapping from the raw PESQ score to the PESQ MOS-LQO score. Modification of PESQ has also
been done to improve performance at estimating quality of speech in a low rate codec of less than 4 kbits/s [13].

PESQ algorithm showed limitations in testing of handsets and other terminals using acoustic interfaces carried out by Rix et al. [13]. In collaboration with the ITU-T Study Group 12 on the development of Acoustics Assessment Model (AAM), they brought about changes to the input filter, equalisation, masking and perceptual model of PESQ. With these, they extended PESQ to create an acoustic model with a wider scope.

The use of objective quality estimation techniques like PESQ in assessing distortions suffered by speech coding and transmission were studied for their ability at predicting the quality of speech enhancement carried out by noise suppression algorithms from the perspectives of signal distortion, noise distortion, and overall quality [14]. This led to the development of a modified version of PESQ in which the weighting coefficient parameters of the linear combination of symmetrical and asymmetrical disturbances for the computation of PESQ score were optimised. It provided for improved quality estimation of transmitted speeches over telecommunication networks.

Focusing on the frame-by-frame time alignment stage of PESQ, [15] noted that subjective scores may be poorly correlated as a result of errors in the objective quality scores caused by a few misaligned frames. Whereas, [16] discovered that PESQ time alignment failed to align continuous variable delays particularly with speech signals that have high packet loss rate and for which dynamic time processing is exhibited due to its piecewise constant delay estimation. The result of Malfait et al’s work achieved a near-perfect delay profile in which for a misalignment of 10 ms, they obtained a correlation of 0.93 with the subjective score, for misalignment less than 5ms they obtained a correlation of 0.973, and have no significant improvement in the correlation coefficient for misalignment down to about 1 ms. They concluded that a time alignment of ±5 ms seemed good enough for correct assessment of time-warped signals. But [16] developed a new time-alignment algorithm that identifies both fix and variable delays in speech signals by using Dynamic Time Warping (DTW) in place of the utterances correlation and splitting methods used in the original PESQ algorithm.

What is known as a New PESQ (NPESQ) was developed by several authors [17,18], based on replacing the auditory perceptual frequency scale, Bark, used in the ITU-T PESQ algorithm with the Equivalent Rectangular Bandwidth (ERB) scale. They claimed the ERB scale is more accurate than the Bark scale for the description of the frequency selectivity of the human auditory system at lower frequencies. They also replaced the Moore and Glasberg loudness model with the Zwicker loudness model. Validating their works on three different wireless codecs, they obtained better correlation coefficients in each case than what was obtained using the normal PESQ.

3. REVIEW OF MAPPING FUNCTIONS

When we run the PESQ algorithm to determine the quality of a degraded speech referenced to the original speech, the raw PESQ quality score within the range -0.5 to 4.5 was mapped to allow for linear comparison with the MOS scale of 1.0 to 5.0 using a mapping function. Versions of mapping functions that have been developed for this purpose were studied and subsequently improved upon in this work.

3.1 The ITU-T Recommended Mapping Function

The ITU-T recommended separate mapping functions for narrowband and wideband speeches. For narrowband speeches ITU-T Rec. P.862.1 [12] was standardised for mapping output of PESQ algorithm and is given by:

\[
    y = 0.999 + \frac{4.999 - 0.999}{1 + e^{-1.4945x+4.6607}}
\]

where, \( x \) is the raw PESQ scores and \( y \) is the mapped PESQ score given as PESQ Mean Opinion Score Listening Objective Quality (PESQ MOS-LQO).

For wideband speech signals (50 – 7,000 Hz), which allows for increased quality and intelligibility, the WB-PESQ mapping function standardised as ITU-T Rec. P.862.2 [19] is given by:

\[
    y = 0.999 + \frac{4.999 - 0.999}{1 + e^{-(-1.668+3.8 x+3.4)}}
\]

This mapping function was developed by simulation data from seven subjective experiments made up of five purely wideband
speech data sets and two narrowband and mixed speech data sets [19].

3.2 The Auryst Mapping Function

The first Auryst’s mapping function, which mapped raw quality score to a dB quality score and then into the MOS score, was noted by [20] to be the first mapping function to be developed. It was further developed by LCC international and purchased by Ericsson. They also noted that Auryst developed a second mapping function, which was a logistic function, given by:

$$y = a + \frac{b - 1}{1 + e^{c x + d}}$$

(3)

where parameters a, b, c, and d are constants optimised for the mapping.

3.3 Morffit and Cotanis Logistic Function

The logistic mapping function developed by Barriac et al. [20] and patented by the United States Patent on Feb 5, 2008, was aimed at achieving improvements in the accuracy of mapping from the raw PESQ scale to the subjective MOS scale. It was given by:

$$y = 1 + \frac{4}{1 + e^{-1.72 x + 5.018}}$$

(4)

This logistic (mapping) function was acclaimed to be more accurate than earlier ones and provided better fit and improvement to the PESQ algorithm performance.

3.4 The Barriac et al Mapping Function

Barriac and his colleagues developed a mapping function in 2004 for use with the PESQ algorithm for wideband signals even before the introduction of the mapping function in ITU-T P.862.2 [21]. The function is given as equation (5) and the plot shown in Fig. 1.

$$y = 1 + \frac{4}{1 + e^{-2x+6}}$$

(5)

3.5 The Sigmoid Curve (S-Curve)

Most of the mapping functions reviewed above are adjusted versions of the logistic population growth function. The logistic growth model is a reliable forecast or prediction model for functional changes, originally developed as a differential equation by Verhauult’s in 1838 [22]. It is represented as a simple sigmoid S-curve given by [23]:

$$\frac{dP}{dt} = r_{max} P \left(1 - \frac{P}{K}\right)$$

(6)

where, P is the population size that ultimately grows to the carrying capacity, K, at time infinity, and $r_{max}$ is the maximum growth rate which occurs at the point of inflection where exponential growth stops and growth or functional change continues as bounded exponential growth. The carrying capacity, K, is actually a point of saturation or stability of the population, while $\left(1 - \frac{P}{K}\right)$ is the fractional deficiency of the instantaneous population function from the peak, K.

![Fig. 1. The Barriac et al mapping function](image-url)
4. METHODOLOGY

Partial integration of the logistic population growth equation produced the solution given by:

\[ P = \frac{Ke^{rt+c}}{1 + e^{(rt+c)}} = \frac{K}{1 + Ce^{-rt}} \quad (7) \]

where, \( C = e^{-c} \) is a constant coefficient.

Replacing the function \( P \) with \( y \) and time, \( t \) with arbitrary variable, \( x \), the solution becomes:

\[ y(x) = \frac{K}{1 + Ce^{-rx}} \quad (8) \]

Adopting a four-parameter approach consisting of coefficients: \( a, b, c \) and \( d \), the function shown in Fig. 2 was developed. It provides detailed description for the determination of the range of steepness of the Sigmoid curve and the \( x \) and \( y \) offsets of the logistic function. This becomes particularly important because none of MOS scale or raw PESQ scores starts from the zero point.

Parameter \( a \) is the full range of the growth function of \( y \) with parameter \( d \) being the offset from the origin. Parameter \( d \) stands for the minimum vertical value. Parameter \( b \) determines the steepness of the curve, and parameter \( c \) determines the midpoint value of the curve.

The \( x \)-axis represents the raw PESQ range between -0.5 and 4.5 while the Subjective MOS range between 1.0 and 5.0 lies on the \( y \)-axis. The offsets on \( x \) and \( y \) axes are such that the initial condition of the logistic function, \( y(x_0) \), which is not necessarily the same as \( y(x_1) \) because of the offset, requires the function be rewritten as:

\[ y(x) = y(x_1) + \frac{K - y(x_1)}{1 + Ce^{-rx}} \quad (9) \]

In Fig. 2, \( y(x_1) = d = 1 \) is the offset on the \( y \)-axis, while on the \( x \)-axis, the offset is 0.5. Actual carrying capacity \( K = a + d = 5 \), and the point of inflection is:

\[ \left[ \ln \frac{C}{r}, \frac{a}{r} + d \right] \quad (10) \]

The function for the mapping was re-written as:

\[ y(x) = 1 + \frac{4}{1 + e^{-(rx+c)}} \quad (11) \]

5. RESULTS ANALYSIS

Efforts at optimising parameters \( b \) and \( c \), of the logistic mapping function:

\[ y(x) = 1 + \frac{4}{1 + e^{-((ax+c)}} \quad (12) \]

took into consideration the boundary and range conditions stated below:

\[ x: \quad \text{Scale of raw PESQ model results:} \quad -0.5 \text{ to } 4.5 \]
\[ y(x): \quad \text{Scale of ideal Subjective MOS scores:} \quad 1.0 \text{ to } 5.0 \]

Case 1:

Substituting the minimum range values, \((x, y) = (-0.5, 1.0)\), into the function results in an error, that is:

\[ \frac{4}{1 + e^{-(bx+c)}} = 0 \quad (13) \]

Case 2:

Substituting the maximum range values, \((x, y) = (4.5, 5.0)\), into the function also results in an error, because \( e^{-(bx+c)} = 0 \). This is because, \( y \) is never equal to 5.0 at \( x = 4.5 \) except at \( x = \infty \) according to the limit:

\[ \lim_{x \to \infty} y(x) = 5.0 \]
Therefore, the following conditions were adopted. With considerations on these boundary conditions and the need to approximate the Sigmoid rule of symmetry about the point of inflection, a number of assumptions were taken.

Condition 1:
Taking off 0.005 from the margin at the bottom and the top of the range resulted in parameters: \( b = 2.6733 \) and \( c = -5.3467 \), and the function becomes:

\[
y(x) = 1 + \frac{4}{1 + e^{-2.67 \times 3.853467}}
\]  

(14)

Condition 2:
Taking off 0.01 from the bottom and the top boundary value resulted in \( b = 2.9563 \) and \( c = -4.7912 \), and the function becomes:

\[
y(x) = 1 + \frac{4}{1 + e^{-2.9563 \times 4.7912}}
\]  

(15)

Going slightly away from exact symmetry or point of inflection of the Sigmoid curve’s, and choosing narrow margins both at the bottom and at the top of the scale of \( y(x) \) to enhance the shape of the resultant curve, the following functions were obtained as indicated on Table 1:

Conditions 3 & 5:
\[
y(x) = 1 + \frac{4}{1 + e^{-2.53 \times 4.5161}}
\]  

(16)

Condition 4:
\[
y(x) = 1 + \frac{4}{1 + e^{-2.22 \times 5.5781}}
\]  

(17)

Condition 6:
\[
y(x) = 1 + \frac{4}{1 + e^{-2.67 \times 4.9531}}
\]  

(18)

MATLAB plot of the logistic function in equations (14) to (18) is shown in Fig. 3. Fig. 4 is the plot obtained when these functions are compared with existing standard mapping functions, the ITU-T Rec. P.862.1 and the United States patented Morfitt and Cotanis logistic mapping function.

### 6. CHOOSING THE PROPOSED LOGISTIC MAPPING FUNCTION

Correlation coefficients were calculated for the test quality scores of transmitted speeches using the logistic functions obtained from the six conditions in section 5. The results are given in Table 2. From the table, the logistic function for condition 4 is best correlated to the subjective MOS with a correlation coefficient of 0.849. The logistic function for condition 4 was therefore chosen as the proposed improved function.

The proposed logistic function is shown in Fig. 5. It was compared, in terms of the coverage of the MOS scale, with the two known standard mapping functions, namely: the ITU-T Rec P.862.1 and the United States patented Morfitt and Cotanis logistic mapping function. The results of the comparison are stated in Table 3 and the plot of the three functions is shown in Fig. 6.

The proposed logistic mapping function achieved 98.6% coverage of the range of the MOS quality score scale. This is a better coverage over those of the existing known standard mapping functions, whereby the ITU-T Rec. P.862.1 function has 86.8% and the Morfitt and Cotanis function has 93.7% coverage of the MOS range. This is an improvement of 11.8% over ITU-T P.862.1 and 4.9% over Morfitt and Cotanis mapping functions respectively.
Table 1. Parameters for the logistic conditions and functions

<table>
<thead>
<tr>
<th>Conditions</th>
<th>Adjustment to lower boundary</th>
<th>Optimised parameters</th>
<th>Remark</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Add</td>
<td>Minus</td>
<td>b</td>
</tr>
<tr>
<td>1</td>
<td>0.005</td>
<td>0.005</td>
<td>4.995</td>
</tr>
<tr>
<td>2</td>
<td>0.010</td>
<td>0.010</td>
<td>4.990</td>
</tr>
<tr>
<td>3</td>
<td>0.005</td>
<td>0.010</td>
<td>4.990</td>
</tr>
<tr>
<td>4</td>
<td>0.005</td>
<td>0.050</td>
<td>4.950</td>
</tr>
<tr>
<td>5</td>
<td>0.010</td>
<td>0.005</td>
<td>4.995</td>
</tr>
<tr>
<td>6</td>
<td>0.010</td>
<td>0.050</td>
<td>4.950</td>
</tr>
</tbody>
</table>

Fig. 4. Comparison of conditioned logistic functions with existing ones
7. TESTING THE MAPPING FUNCTIONS

Putting the proposed logistic mapping function side-by-side the two standard functions with which it was compared, the variability in their mapped data were evaluated using analysis of variance (ANOVA). At a significance level of $\alpha = 0.05$, results obtained were F-statistical value of 60.6042, and a critical-F of 3.04, and a $p$-value of $4.6172E-21$. With $p < 0.05$, the Null Hypothesis was rejected, and the critical-F value being less than the F-statistic value confirmed the rejection. Therefore, the data distribution of at least one of the functions has a different mean and belongs to a separate population of performance.

### Table 2. Basis on which the proposed condition was chosen

<table>
<thead>
<tr>
<th>Condition</th>
<th>Correlation coefficient</th>
<th>Correlation rating</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.789902</td>
<td>Least correlated</td>
</tr>
<tr>
<td>2</td>
<td>0.806808</td>
<td>Second to the last</td>
</tr>
<tr>
<td>3, 5</td>
<td>0.837423</td>
<td>Third best correlated</td>
</tr>
<tr>
<td>4</td>
<td>0.849006</td>
<td>Best correlated</td>
</tr>
<tr>
<td>6</td>
<td>0.845914</td>
<td>Second best correlated</td>
</tr>
</tbody>
</table>

### Table 3. Comparing proposed mapping function with two prominent functions

<table>
<thead>
<tr>
<th>S/N</th>
<th>Raw PESQ score</th>
<th>Subjective MOS score</th>
<th>ITU-T Rec. P.862.1 mapped PESQ MOS score</th>
<th>U. S. Patented logistic function mapped PESQ MOS score</th>
<th>Proposed logistic function mapped PESQ MOS</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>-0.5</td>
<td>1</td>
<td>1.077321721</td>
<td>1.011137984</td>
<td>1.00499980</td>
</tr>
<tr>
<td>2.</td>
<td>4.5</td>
<td>5</td>
<td>4.548638319</td>
<td>4.757634956</td>
<td>4.95000751</td>
</tr>
<tr>
<td>Difference between highest &amp; lowest scores</td>
<td>4</td>
<td>3.471316598</td>
<td>3.746496972</td>
<td>3.94500771</td>
<td></td>
</tr>
<tr>
<td>% of MOS Score</td>
<td>100%</td>
<td>86.8% of MOS</td>
<td>93.7% of MOS</td>
<td>98.6% of MOS</td>
<td></td>
</tr>
</tbody>
</table>

**Fig. 5. Plot of proposed logistic function**
8. CONCLUSION

Continuous improvements in the techniques and models developed and adopted in testing for and in the estimation of the quality of processed and/or transmitted speech signals have been ongoing in the last two to three decades. Part of these efforts which led to the development of the state-of-the-art PESQ estimation model for objective quality testing has also been made to undergo a number of constraints and limitation evaluations leading to improvements in major aspects of its functions. This work evaluated and improved on the mapping function of PESQ model with the development of a logistic function that boast of 98.6% coverage of the subjective MOS scale. It thereby provides better merit of fitness to the subjective MOS which is the basis of correlation of all objective quality estimation techniques and models.

COMPETING INTERESTS

Authors have declared that no competing interests exist.

REFERENCES


