

# NIDA: A Parametric Vocal Quality Assessment Algorithm over Transient Connections

Sofiene JELASSI<sup>1,3</sup>, Habib YOUSSEF<sup>1</sup>, Lingfen SUN<sup>2</sup>, and Guy PUJOLLE<sup>3</sup>

<sup>1</sup> Research Unit PRINCE, ISITCom, Hammam Sousse, Tunisia

<sup>2</sup>School of Computing, Communications and Electronics, University of Plymouth, UK

<sup>3</sup>Laboratory of Computer Science (LIP6), University of Pierre and Marie Curie, Paris, France

Sofiene.Jelassi@infcom.rnu.tn, habib.youssef@rnu.fsm.tn,

l.sun@plymouth.ac.uk, and Guy.Pujolle@lip6.fr

**Abstract.** This paper presents NIDA, a Non-Intrusive Disconnection-aware vocal quality assessment Algorithm. NIDA accurately estimates vocal perceived quality over wireless data networks by discriminating the perceptual effect of a single random packet loss, 2-4 consecutive packet losses (burst) stemming from contentions, and discontinuity entailed by transient loss of connectivity. NIDA properly accounts for transient loss of connectivity experienced by mobile users over wireless data networks, stemming from vertical and horizontal handovers, or when users roam out of the coverage area of the associated infrastructure. To this end, a novel lossy wireless data channel model has been conceived based on a continuous-time Markov model. The channel model is calibrated at run-time based on a set of measurements gathered at packet layer using the header content of received voice packets. The perceived quality under each state is properly quantified, then combined in order to predict quality degradation due to wireless data channel features. Performance evaluation study shows that quality degradation ratings calculated using NIDA strongly correlate with quality degradation ratings calculated based on ITU-T PESQ intrusive algorithm, which mimics tightly subjective human rating behavior.

**Keywords:** Perceptual voice quality, Transient wireless connections, E-model.

## 1. Introduction

In recent years, Voice over IP networks, denoted as VoIP, becomes a popular service. Since its inception, huge strides have been made. Currently, VoIP has an increasing widespread popularity and used as alternative to traditional telephony in homes and enterprises. Unlike circuit-switched telephone networks, ordinary IP networks cater to applications a best effort service resulting in packet loss and variable network delay. These features significantly harm perceived quality of delay-sensitive service such as VoIP. These sources of disturbance negatively influence the perceptual quality in two ways, intelligibility of the speech sequence and interactivity. There are several ways to improve perceived quality over ordinary IP networks. Basically, the quality metric indicators can be improved in network- or application- centric ways. Network-centric approaches improve service quality by adequately upgrading core nodes to accommodate features of delay-sensitive services such as service differentiation and resource reservation [1]. In contrast, application-layer approaches improve service

quality through an intelligent control of delivered media at sender and receiver sides [1]. It is highly desirable to evaluate the suitability of developed QoS mechanisms at users' level. Measuring accurately the perceptual vocal quality is pivotal from operators as well as customers' perspectives. In fact, perceptual quality can be used to rate service providers. Moreover, telecom operators can use assessment algorithms for management, maintenance, monitoring, planning, and diagnosis operations. On the other hand, subscribers can use perceived quality to select adequate access network under a given circumstance. Indeed, services over next-generation networks will likely be offered to users using a multitude of overlapping networks and terminals. In such a case, subscribers can select the configuration that responds to their preferences.

The assessment of voice quality at the user level can be performed subjectively or objectively. Subjective-based approaches derive the vocal quality using a set of human subjects which vote perceived quality under a given situation [2]. The vocal quality assessment is performed using a dedicated scale. The ITU has defined a standard subjective metric called the Mean Opinion Score (MOS) to quantify listing quality under ACR (Absolute Category Rating) subjective tests [2]. MOS scores vary from 1 (bad quality) to 5 (excellent quality). Certainly, subjective approaches are unable to rate at run-time the perceptual quality of live vocal conversations which confine their utility to a limited range of applications. Moreover, subjective approaches are time consuming, cumbersome, and expensive.

Rather than using human subjects to rate vocal quality, objective-based approaches estimate perceived quality using machine-executable algorithms running either on end- or mid- nodes [3]. There are several assessment algorithms reported in the literature which can be classed as *signal layer black box* and *parametric model glass box* categories [3]. Signal layer black box algorithms estimate the perceived quality by properly processing speech signals without knowledge of the underlying transport network and terminals features. In contrast, parametric model glass box algorithms require full characterisation of transport network and terminals to estimate perceived quality. Technically, parametric assessment algorithms are more attractive, especially over packet-based networks, due to their reduced complexity and their suitability to manage and monitor live packetized vocal service. This is made without acceding to speech sequences which is desirable for security reasons. However, parametric assessment algorithms are relatively less accurate than signal layer black box assessment algorithms.

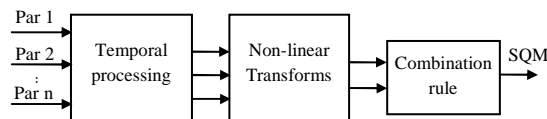
In order to accurately estimate perceived quality based on parametric model paradigm, there is a need of rigorously defining appropriate input metrics, models, and combination rules. The ITU-T has standardized a parametric model-based assessment algorithm, called E-Model, which requires full characterization of underlying system (network and terminals) to estimate/predict perceived quality [4]. Suitable parameters, models, and combination rules have been defined and calibrated based on extensive conducted subjective experiences. ITU-T E-Model has been initially developed to evaluate perceived quality over wired Telecom networks. Since then, several revisions have been made by academics and industrials in order to increase its accuracy over a wide range of networks, especially over packet-based and wireless networks [4].

The main goal of this work is to properly build adequate objective models which are able to model and quantify the distortions encountered by VoIP-interlocutors over shared wireless data links. Besides traditional impairments incurred by users over packet-based networks, VoIP-customers over wireless data networks experience transient loss of connectivity stemming from either vertical or horizontal handovers, or when mobile users roam out of the coverage area of the associated network. It is well-recognized that such a disturbing event should be properly accounted for in the computation of perceived quality [5]. To properly include transient losses of connectivity in the vocal quality estimate, a novel lossy wireless channel model has been conceived based on a continuous-time Markov model. The conceived model incorporates three states which stand for a single random loss, 2-4 consecutive packet losses (a burst), and discontinuity. The channel model is calibrated at run-time using a set of network measurements gathered at packet layer. The perceived quality at the end of each assessment period is calculated through the combination of perceived quality in each state. To improve accuracy, the temporal location of transient loss of connectivity within an assessment period is considered in the calculation of the overall perceived quality. Performance evaluation study shows that our perceptual vocal quality algorithm, NIDA, produces well-correlated scores with ITU-T PESQ intrusive algorithm. Note here that conducting formal subjective MOS tests on a large scale is beyond any reasonable allocated time and budget. The ITU-T PESQ full-reference vocal assessment algorithm models accurately the subjective human rating behavior [3].

The remainder of this paper is organized as follows. Section 2 describes the principles of parametric models. Section 3 presents the novel wireless data channel model used to account for the different types of loss-related disturbances. In Section 4, we present how perceived quality is calculated under each state and how the overall perceived quality is produced. In Section 5, we compare the performance of NIDA against the ITU-T PESQ intrusive algorithm. We conclude in Section 6.

## 2. Principle of parametric perceptual models

Parametric model glass box assessment algorithms estimate subjective speech quality using a set of input parameters gathered from the network and edge-devices. As illustrated in Figure 1, input parameters can be transformed to increase their correlation with subjective scores. Then, a combination rule, called also perceptual model, is used to estimate the Speech Quality Measure (SQM) metric, e.g., MOS.



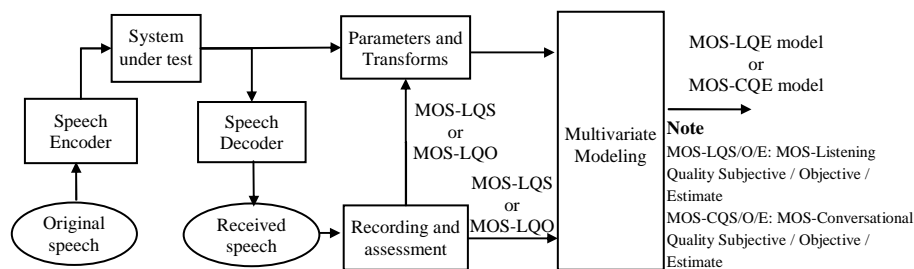
**Figure 1:** Principle of parametric SQM, input parameters, and basic processing steps.

As we can deduce, selection of pertinent input parameters, transformations, and perceptual models should be performed off-line. To this end, dedicated assessment frameworks for vocal quality modeling should be set-up. As illustrated in Figure 2, a large set of speech samples are processed by a system under test to produce degraded speech sequences under a given circumstance. The corresponding parameters for full system characterisation are properly recorded. The degraded speech sequences are evaluated either by human subjects or a signal layer assessment algorithm, which is deemed sufficiently accurate to mimic human behavior rating such as ITU-T PESQ. The system under test can be either a simulated or emulated voice transport system or an experimental test-bed.

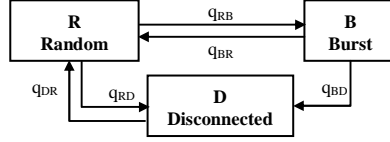
Pertinent input parameters for full system characterization such as mean packet loss rate, loss pattern, delay, delay jitter, echo, side-tone, coding, packet loss concealment, and noises are closely dependent on the delivering systems and used terminals. In this work, we study packet-based vocal conversation over infrastructure-based shared wireless data networks such as IEEE 802.11, WiMAX, and Wi-Fi. The relevant sources of disturbance observed over data networks are packet loss, delay, and delay jitter. Moreover, over wireless networks, moving interlocutors encounter vertical and horizontal handovers, which are made to improve service quality and to assure service ubiquity.

In the context of ITU-T E-Model methodology, the effect of delay impairment factor ( $I_d$ ) on perceptual quality, which influences interlocutors' interactivity, has been rigorously modeled in the literature [6]. In contrast, the effect of equipment impairment factor ( $I_e$ ) on perceptual quality, which influences speech intelligibility, remains unsatisfactory under several situations [7]. This is due to the diversity of disturbing sources which include, among others, packet loss, de-jittering buffer management, coding switching, tandem configuration, and handovers. Moreover, rapid evolution of access and core networks as well as terminals entails complication in the development of accurate perceptual models. That is why  $I_e$  models need to be re-configurable and flexible as much as possible.

The main contribution of this work is the proposal of an adequate objective model that estimates the equipment impairment factor over a shared wireless data link. We are unaware of any similar work in the literature which targets the same goal.



**Figure 2:** Vocal assessment framework for non-intrusive quality modeling.



**Figure 3:** Markov process model of packet losses over a wireless data channel

### 3. Loss modeling over wireless data networks

It is well-recognized that packet loss behavior over cabled IP networks is bursty. That is why, a 2-state Markov model has been widely used in the literature in order to model and analyze packet losses over wide area IP networks [8]. The bursty loss behavior stems mainly from network congestion which still happens over shared wireless data networks. However, during certain periods of time, packet-based mobile receiver will likely incur random packet losses due to signal-related problems such as fading and interferences. In fact, in such a circumstance, link layer protocol attempts to recover lost voice packets through retransmissions which will entail additional transmission delays. In such a case, delayed voice packets will likely be ignored by the play-out process because they reach the receiver side after their playout instants. Moreover, mobile receiver incurs, during certain periods of time, transient losses of connectivity during handovers or when users roam out of coverage area of associated access point. It is important here to highlight the perceptual difference between a packet burst loss which lasts typically for 2-4 voice packets and disconnection which means that users will clearly hear a discontinuity in the rendered stream. In fact, a burst of 2-4 voice packets can be efficiently recovered by modern CODECs and users hearing brain [3].

To precisely account for wireless data networks behavior from users' perspective, we propose to model a wireless channel using a continuous-time Markov process (see Figure 3). The modeled stochastic process takes its values from the following 3-dimension state space: R (for random), B (for burst), and D (for disconnected). When in state R, packet losses are generated randomly according to a Bernoulli distribution. This behavior guaranties that packet losses occur in uncorrelated way for a given PLR (Packet Loss Ratio). However, when in state B, packet losses are generated in burst according to the traditional Gilbert/Elliot model [9]. The burstiness of packet losses is controlled using the ULP (Unconditional Loss Probability) and CLP (Conditional Loss Probability). When at state D, all sent packets are lost.  $q_{IJ}$ , where  $I$  and  $J \in \{R, B, D\}$ , represents the transition rate from state  $I$  to  $J$ .

As illustrated in Figure 3, after a residence period in state B, the process will enter state R with probability  $q_{BD}/(q_{BR} + q_{BD})$  and state D with probability  $q_{BR}/(q_{BR} + q_{BD})$ . The residence periods in states R, B, and D follow an exponential distribution with mean values equal to  $T_R$ ,  $T_B$ , and  $T_D$  respectively. The mean residence time in each state can be computed as follows:

$$T_R = 1/(q_{RB} + q_{RD}) \quad T_B = 1/(q_{BD} + q_{BR}) \quad T_D = 1/q_{DR} \quad (1)$$

#### 4. Parametric perceptual models over shared wireless data channel

The main goal of parametric single-ended perceptual models consists of providing timely feedbacks about perceived quality at run-time of a live vocal session. The estimated perceived quality is included in QoS reports sent periodically to the sender or the policy enforcer nodes. For the sake of monitoring, the recommended assessment window lasts between 8 and 20 seconds [6]. The VoIP receiver gathers suitable measurements which can be transformed then combined to estimate perceived quality.

In order to develop suitable parametric perceptual quality estimate models, the modeling framework depicted in Figure 2 should be set-up. In our case, the impairments introduced by the system under test are rigorously modeled as in Figure 3. The level of introduced impairments can be finely calibrated according to the mean residence time and loss parameters in each state. In order to avoid useless extensive experiments which have been already made in the literature, we propose the following strategy:

- In states R, B and D, we use perceptual speech quality estimate models available in the literature.
- Perceived qualities under each state are meticulously combined to produce overall perceived quality at the end of an assessment period.

Surely, when perceptual models under R or B state are unavailable, the modeling framework described in Figure 3 can be used to develop such quality estimate models. In order to clearly illustrate our methodology, we give as guideline how to develop a perceptual model of ITU-T G.729 speech CODEC over wireless data channel. The G.729 CODEC is recommended over a wide range of configurations, especially over reduced capacity and lossy wireless channel [9, 10]. From [6], it has been shown that the equipment factor under random packet loss,  $I_{e-R}$ , is given by:

$$I_{e-R}(G.729, plr) = 11 + 40 \times \ln(1 + 10 \times plr) \quad (2)$$

where,  $plr$  represents the mean packet loss ratio encountered by the receiver during a random loss period. Note that (2) includes disturbances due to G.729 speech CODEC and mean packet loss ratio. In fact, distortion stemming from coding scheme can be obtained for a packet loss ratio set to 0, which is equal to 11 for G.729.

When in burst loss state, we use perceptual models presented in [11]. Authors identify loss pattern and degree of burstiness by recording inter-loss gaps preceding loss bursts, in a series of (gap, burst) pairs [11]. The perceptual effect of each single pair is estimated using a perceptual model which accepts as input the gap and burst lengths expressed in packets. The perceived quality at the end of an assessment interval is derived through a weighted aggregation of produced scores. Specifically, authors show that the following expression accurately estimates speech quality:

$$\text{MOS}_B(\{\text{gap}_i, \text{burst}_i\} / 1 \leq i \leq P) = \frac{\sum_{i=1}^P (\text{gap}_i / 10) \times \text{MOS}_i(\text{gap}_i, \text{burst}_i)}{\sum_{i=1}^P (\text{gap}_i / 10)} \quad (3)$$

where,  $\text{gap}_i$  and  $\text{burst}_i$  are, respectively, the length of inter-loss and loss durations of  $i^{\text{th}}$  (gap, burst) pair,  $\text{MOS}_i$  is the “base” quality model used for  $i^{\text{th}}$  (gap, burst) pair, and  $P$  represents the number of (gap, burst) pairs observed in burst state. Once  $\text{MOS}_B$  score is estimated, the equipment impairment factor under burst loss,  $I_{e-B}$ , can be calculated as follows:

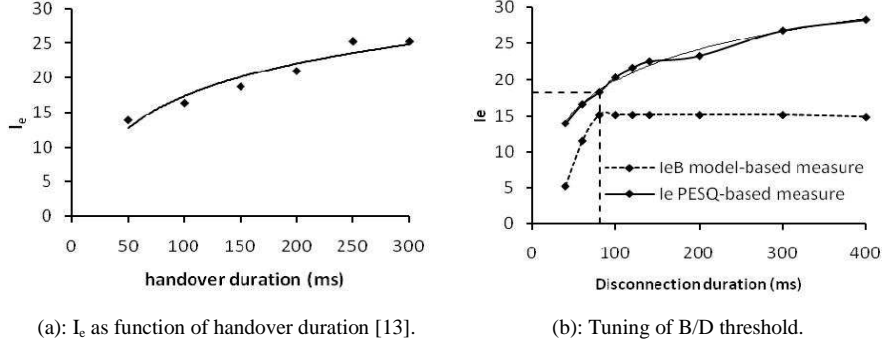
$$I_{e-B}(G.729) = 93.2 - \text{MOS2R}(\text{MOS}_B(\{\text{gap}_i, \text{burst}_i\})) \quad (4)$$

where,  $\text{MOS2R}$  refers to the function which allows converting a quality score from the MOS domain to the rating factor domain [6, 8, 9]. The transient loss of connectivity significantly impairs the quality of users’ experience. In fact, loss of connectivity entails service discontinuity which greatly degrades perceptual quality. In fact, beyond a certain threshold, such a temporary discontinuity will lead to the abrupt hang-up of voice sessions. Basically, service interruptions are entailed by horizontal (intra-) and vertical (inter-) network handovers. Typically, the procedure “make-before-break” is used during handovers for delay-sensitive services which reduces significantly the latency to change associated access point (AP). However, handover delay can be dependent on the actual cell load, AP search procedure, and authentication mechanism. Moreover, handover between Inter-network domains needs looking after IP address which could increase handover latency [12].

In [13], A. F. Duran et al. studied the effect of handover over wireless data networks on perceived quality. Important results are presented in the curve plotted in Figure 4a which shows the equipment distortion factor as a function of handover duration. In order to quantify the effect of handover on perceived quality at run-time, we build the following quality estimate model based on a logarithmic regression process applied on the set of obtained subjective scores:

$$I_{e-D}(G.729, T_D) = (6.1913 \times \ln(T_D) - 8.6216) \times L \quad (5)$$

where,  $T_D$  represents the handover/disconnection duration. The quality estimate model achieves a square correlation factor equal to 0.98. Note that discontinuity is only considered during active periods, where it influences perceived quality unlike silence periods. The coefficient  $L$  is a weighting factor used to account for handover location. In fact, earlier studies have shown that pertinent disturbing events occurring close to the end of a voice conversation disrupt more negatively users’ experience [14]. According to a set of extensive subjective experiences made by France Telecom, we assign to  $L$  the value of 1, 0.9, and 0.78 when a handover occurs, respectively, at the end, middle, and beginning of an assessment period [14].



**Figure 4** : Discrimination between effect of 2-4 packet loss and handover duration for G.729B speech CODEC.

In Figure 4b, we plot speech quality as a function of disconnection duration introduced in the middle of a set of examined sequences. The speech quality is measured based on ITU-PESQ algorithm and predicted based on  $I_{e-B}$  model proposed in [11] using the framework depicted in Figure 2. These curves show clearly that  $I_{e-B}$  model is unable to accurately estimate the perceptual effect of disconnection. As we can see, the obtained results are well-correlated with subjective trials performed in [13]. Further details regarding empirical trials will be given later in the evaluation section. We experimented with several expressions to quantify the overall service quality degradation over wireless data channels at the end of a speech assessment period. Based on preliminary experimental results, the following model has been selected:

$$I_e(av) = \alpha_0 + \alpha_1 W + \alpha_2 W^2 \quad \text{where} \quad W = \frac{T_R \times I_{e-R} + T_B \times I_{e-B} + T_D \times I_{e-D}}{T_R + T_B + T_D} \quad (6)$$

where,  $W$  corresponds to the experienced average equipment factor over time, and  $\alpha_0$ ,  $\alpha_1$ , and  $\alpha_2$  are fitting coefficients obtained using polynomial regression. A series of equipment values are produced during a vocal conversation, which are averaged over time to quantify perceived quality at the end of service using ITU-T E-Model as follows [8]:

$$R = 93.2 - I_{e\text{-weighted}}(av) - I_{d\text{-weighted}} \quad (7)$$

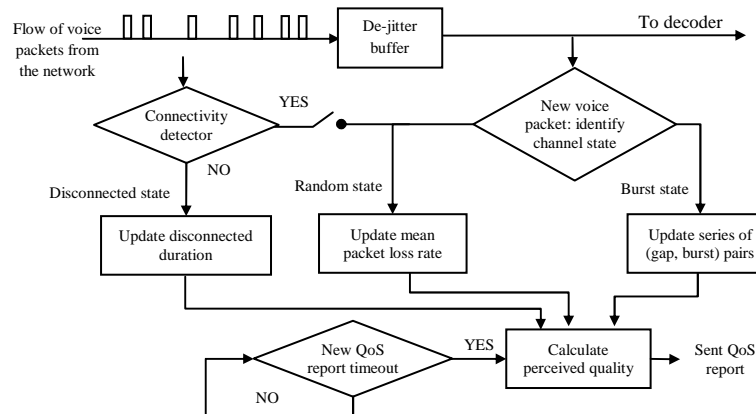
where,  $R$  is the rating factor varying between 0 (worst quality) and 100 (excellent quality),  $I_{e\text{-weighted}}$  and  $I_{d\text{-weighted}}$  represent, respectively, the weighted average over time of distortions due to equipment and delay. According to empirical subjective experiences, the mean of instantaneous perceived quality correlates well with subjective opinion scores given by humans at the end of a voice conversation [14].



## 5. Architecture of the vocal quality assessment NIDA

As mentioned earlier, the developed vocal assessment tool is intended to evaluate voice service over transient connections at run-time of live voice conversations. This is performed by examining the header content of each received packet. As illustrated in Figure 5, our assessment approach, NIDA, examines received packets before and after de-jittering buffer. This allows, on the one hand, accounting for ignored packets at the de-jitter buffer, and on the other hand, reliably identifying channel connection state which is determined using a passive connectivity detector.

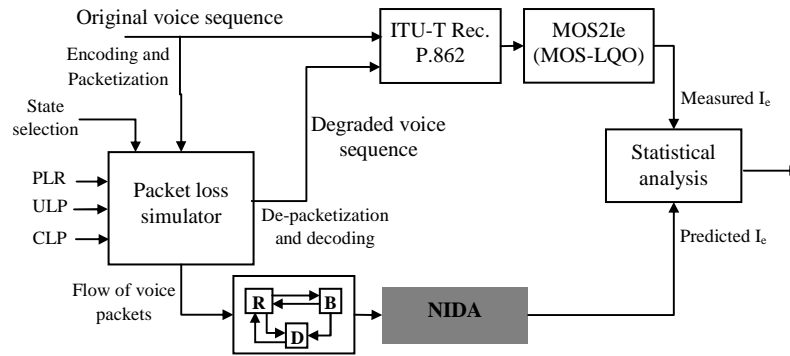
The packet loss process is only accounted for when communicating terminals are connected (see Figure 5). In such a case, the assessment voice algorithm classifies eventual lost packets under burst or random states as follows. If several successive voice packets are lost, then missing segments are accounted for in burst state. In such a case, NIDA updates the series of (gap, burst) pairs (see Figure 5). The value of gap corresponds to the number of consecutive played voice packets between last and current loss instances. The value of burst corresponds to the number of consecutive lost packets of the current loss instance. Note that gap and burst values are calculated according to the sequence numbers of examined packets. If a single packet loss happens, then NIDA checks the number of played packets before the loss occurrence. Then, it classes the loss as random if the gap is greater than  $g_{min}$ , otherwise, it is classed as burst. This is made to consider frequent and temporally-close loss instances as burst loss. In random loss state, the mean packet loss rate and random period duration are updated.



**Figure 5:** Functional diagram of vocal quality assessment NIDA.

The connectivity detector probes passively received voice packets in order to reliably identify channel connection state. The connectivity detector checks the sequence number of each in-sequence received packet. Indeed, out-of-order packets are seldom observed over infrastructure-based networks and can stem mainly from route switching. In contrast to play-out process, late voice packets are considered by

connection detector process. In reality, delay can stem from congestion or reduced data rate switching when wireless interfaces enable multi-rate functionality. The temporary loss of connectivity is decided based on an empiric selected handover threshold. Specifically, we handle a loss instance in disconnection state when the loss duration is greater than 100 ms which corresponds to five 20-ms successive voice packets. Loss durations less than 80 ms are accounted for in burst state.



**Figure 6:** Evaluation framework of NIDA

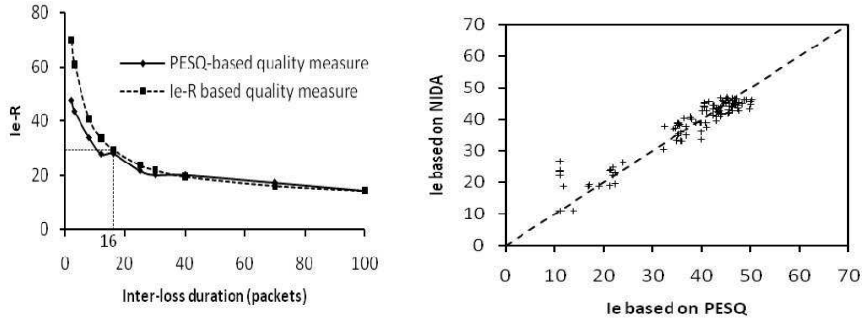
In order to detect the temporal location of a disconnected period in an assessment period, the lower and upper timestamp bounds of disconnected periods are recorded. In reality, it is likely to observe at most one disconnection instance during an assessment period, given the coverage range and human walking speed.

## 6. Calibration and validation of NIDA

A set of empirical trials have been conducted in order to calibrate and validate the suitability of NIDA to evaluate voice conversations over transient connections. To do that, we have developed the quality framework depicted in Figure 6. Actually, packet losses are generated according to channel model presented in Figure 3. The pertinent parameters of conceived packet loss model are the mean sojourn duration in each state which follows an exponential distribution. Moreover, the loss process parameters in random and burst states are given by users. The disturbance stemming from packet loss is measured based on the intrusive signal layer ITU-T PESQ algorithm. On the other hand, the flow of voice packets is examined by NIDA to predict perceived quality. In order to evaluate the accuracy of NIDA, the measured and predicted disturbances are statistically analysed in term of their degree of correlation and the Root Mean-Squared Error (RMSE).

The first series of trials aim at fine-tuning the parameter  $g_{\min}$  used by NIDA to discriminate between random and bursty loss periods. This is done by introducing periodically single packet loss events to sixteen speech sequences spoken by eight male and eight female English speakers, taken from ITU-T P.Supp23 dataset. The inter-loss gap,  $g_{\min}$ , was varied from 3 to 100 20-ms voice packets. The disturbance is

measured using  $I_{e-R}$  speech quality model given in Equation (2). Figure 7a illustrates that a decrease of  $g_{\min}$ , which induces an increase of burstiness, entails a reduction of the  $I_{e-R}$  model accuracy. This observation is somehow expected since the used model is only able to quantify the effect of random loss. According to our empirical measurements shown in Figure 7a, we set the value of  $g_{\min}$  to sixteen 20ms-voice packets (320ms).



(a): Tuning of  $g_{\min}$  value. (b): Scatter-plot of  $I_e$  calculated based on PESQ and NIDA.

**Figure 7:** Calibration and validation of NIDA.

The second series of empirical trials is done to develop/calibrate and validate NIDA. The calibration is performed using the previous dataset, where speech sequences are distorted according to parameters summarized in Table I. The validation dataset contains eight standard ITU-T 8s-speech samples, not used in the training dataset, spoken by four male and female English speakers. The mean duration of Random and Burst periods are set to 2 sec.

**Table I:** Experienced empirical trials to calibrate and validate NIDA.

State	Loss parameter	Level		Cardinality
		Modeling	Validation	
Random	$ULP_R^{(1)}$	0.02	0.03	1
Burst	$ULP_B^{(2)}$	0.15; 0.25	0.10; 0.20	2
	$CLP^{(3)}$	0.20; 0.50; 0.90	0.30; 0.60; 0.95	3
Disconnected	Mean (ms)	150; 250	120; 200	2
Total number of <i>scenarios</i>		$(1 \times 2 \times 3 \times 2) \times 2$		24
Speech material		<i>Training dataset</i> 8 females, 8 males	<i>Validation dataset</i> 4 females, 4 males	
Number of measurements		192	96	288

<sup>(1)</sup>  $ULP_R$ : Unconditional Loss Probability in random state

<sup>(2)</sup>  $ULP_B$ : Unconditional Loss Probability in burst state

<sup>(3)</sup>  $CLP$ : Conditional Loss Probability in burst state

The results produced by the training dataset are used to derive fitting coefficients of the combination rule defined in (6). This statistical analysis indicates that the suitable fitting values are the following:  $\alpha_0 = -17.017$ ,  $\alpha_1 = 2.197$  and  $\alpha_2 = -0.02$ . The

obtained model is used to predict equipment impairment factor of the validation dataset using NIDA (see Table I). Figure 7b represents a scatter-plot showing the relationship between  $I_e$  values measured using ITU-T PESQ and predicted using NIDA for the validation dataset. This plot shows strong correlation between NIDA estimates and PESQ-based intrusive scores. Indeed, we found a correlation factor equal to 0.95 coupled with a Root Mean Square Error of 0.07. Finally, we note the presence of some outliers which deviate from the angle  $45^\circ$ . This deviation is located at loss region characterized by small and random loss behavior. In such a case, the effect of coding, which deviates from one sample to another according to the speech content, significantly influences the overall measured disturbance. Overall, NIDA exhibited excellent accuracy in evaluating, on a per-call basis, voice sequences with bursty losses and transient disconnections.

## 7. Conclusion and future work

This paper introduced NIDA, a Non-Intrusive Disconnection Aware vocal assessment algorithm. NIDA is intended to evaluate vocal quality over channels characterized by a transient loss of connectivity. To do that, a novel data channel model has been conceived based on a 3-state continuous-time Markov process. The perceived quality is quantified at run-time in each state, then properly combined at the end of an assessment period. NIDA discriminates between burst and disconnected periods in the calculation of perceived quality. Evaluation study showed that predicted measures produced by NIDA strongly correlate with ratings given by ITU-T PESQ ( $R = 0.95$ ). As such, the work has extended the current E-model for voice over wireless applications with a consideration of possible voice discontinuity during handover. As future work, we plan to model and evaluate more precisely the effect of large transient disconnection periods on perceptual quality. Moreover, we envisage increasing the accuracy of NIDA by including voicing features of handled voice frames.

## References

1. Melvin, H.: The use of synchronized time in voice over Internet Protocol (VoIP) applications. Ph.D. Thesis, University College Dublin, Ireland, (2004)
2. ITU-T Recommendation P.800: Methods for Subjective Determination of Transmission Quality. (1996)
3. Rix, A., Beerends, J., Kim, D., Kroon, P., Ghitza, O.: Objective assessment of speech and audio quality: Technology and Applications. IEEE Transactions on Audio, Speech, and Language Processing, Vol. 14, No. 6, pp. 1890:1901, (2006)
4. ITU-T Recommendation G.107: The E-Model a Computational Model for Use in Transmission Planning. (2005)
5. Mobisense project: User Perception of Mobility in NGN. In proceeding of DTAG Workshop QoS and QoE monitoring, Berlin, Germany, (2007)
6. Cole, R.G, Rosenbluth, J. H.: Voice over IP Performance Monitoring. Computer Communication Review, ACM Sigcomm, 31(2), (2001)
7. Takahashi, A., Yoshino, H., Kitawaki, N.: Perceptual QoS assessment technologies for VoIP. IEEE Communication Magazine, Vol. 42, No. 7, pp. 28:34, (2004)

8. Carvalho, L., Mota, E., Aguiar, R., Lima, A.F., de Souza, J.N.: An E-model implementation for speech quality evaluation in VoIP systems. In Proceedings of ISCC 2005, (2005)
9. Hoene, C.: Internet telephony over wireless links, PhD thesis, Technical University of Berlin, Germany, (2005)
10. ITU-T Recommendation G.729: Coding of speech at 8 kbit/s using conjugate-structure algebraic-code-excited linear prediction (CS-ACELP). (2007)
11. Roychoudhuri, L., Al-Shaer, E., Settini, R.: Statistical Measurement Approach for On-line Audio Quality Assessment. In Proceedings of Passive and Active Measurement (PAM'06), (2006)
12. Lakas, A., Boulmalf, M.: Study of the Effect of Mobility Handover on VoIP over WLAN. In proceedings of 3rd International Conference on Innovations on Information Technology. Dubai, UAE (2006)
13. Duran, A. F., Pliego, E. C., Alonso, J. I.: Effects of handover on Voice quality in wireless convergent networks. In proceeding of IEEE Radio and Wireless Symposium 2007, Long Beach, California, USA, (2007)
14. France Telecom: Study the relationship between instantaneous and overall subjective speech quality for time-varying speech sequence: influence of a recency effect. ITU Study Group 12, Contribution D.139, (2000)