# NETWORK RESOURCE ALLOCATION FOR PERCEPTUALLY BASED UNEQUAL PACKET PROTECTION IN VOICE COMMUNICATION

*Steffen Præstholm*[1], *Søren Skak Jensen*[2], *Søren Vang Andersen*[2], *Manohar N. Murthi*[3]

[1] Mobile Phones Development.
BenQ Denmark ApS, Denmark
steffen.praestholm@benq.com

[2] Dept. of Communication Technology
Aalborg University, Denmark
{sje00,sva}@kom.aau.dk

[3] Dept. of Electrical and Computer Engineering
University of Miami, USA
mmurthi@miami.edu

## ABSTRACT

We address the problem of optimizing resource allocation for Perceptually based Unequal Packet Protection (PUPP) in a packet based voice carrying network. For that purpose, we design a novel real-time working Perceptually Based Classifier (PBC) optimizing the assignment of voice packets to either a Premium (Pch) or an Ordinary (Och) transmission Channel with regard to packet perceptual importance. In particular, our PBC is based on Sliding Window optimization (SWO) and implement PESQa, an improved method to real-time estimation of speech quality. Based on this PBC and a Differentiated Service (DS) implementation of the Pch/Och, objective results indicate that 70% premium packet assignments optimizes performance over a broad range of loss scenarios on a bottleneck link. Additionally, packet loss statistics gives a clear indication on criteria for optimizing PUPP Pch/Och.

## 1. INTRODUCTION

Real-time packet based voice transmission, like Voice over IP (VoIP) suffers from quality degradation due to packet losses which are combatted through a combination of receiver-based Packet Loss Concealment (PLC), and proactive schemes like packet protection. Typically, proactive packet protection is effected through Forward Error Correction (FEC) which entails the transmission of redundant information (e.g., parity check packets), or network-based schemes such as Differentiated Services (DS) in which different packets are transmitted over virtual channels with varying priorities.

In allocating error-control resources for either FEC or DS, a VoIP application takes either an equal packet protection (EPP) or an unequal packet protection (UPP) approach. Recognizing the unequal perceptual importance of voice packets, several researchers have explored different methods for implementing perceptually unequal packet protection (PUPP) in which packet protection resources are allocated according to perceptual importance [1, 2, 3, 4]. In general, these previous PUPP methods can be viewed as consisting of a perceptual classification followed by an assignment to a Transmission Channel (Tch) as in Figure 1. Here, a Tch is characterized by the utilization cost and the packet loss probability, both of which are functions of e.g., the amount of bits spent on redundancy for FEC or the number of packets that may be placed in high priority/premium classes for a DS-network. Although the previous approaches have reported very promising results, the basic methods utilized in PUPP can be improved.

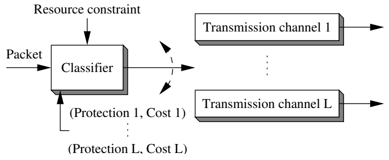

**Fig. 1**. PBC general block-diagram.

For example, previous perceptual classifiers have been primarily based on perceptually simple measures such as spectral distortion possibly augmented by additional distortion measures [3, 1] or simple unvoiced to voiced transitions [4]. Clearly, a perceptually based classifier closer to speech quality measures are warranted.

Moreover, the connection between the PUPP method and the network performance can be more carefully examined. For example, in [1], the effect of packet marking for a two-class (ordinary and premium) DS method for PUPP on the performance of the network is not explored. In [3], it is mentioned that the PUPP-based FEC protection can be allocated such that the sum total of the coding and FEC bits satisfy TCP-friendly rate constraints, but results for such time-varying allocations and network conditions are not presented. In [4], the results are primarily given for a special active queue management scheme. Even in a network containing homogeneous VoIP traffic (which is common for many small commercial VoIP services), the effect of PUPP on network performance has not been fully explored. It is clear that a VoIP application cannot protect all its packets with impunity. As more packets are protected, congestion can increase. Therefore, a PUPP scheme must be flexible to change its allocation policies according to time-varying network conditions. In general, the relationship between PUPP, voice quality, and network performance is of interest.

In this paper, we develop a PUPP scheme that is more perceptually sound and study its effect on both voice quality and overall network performance. In particular, we propose a Perceptually Based Classifier (PBC) whose perceptual importance measure is based on the Perceptual Evaluation of Speech Quality (PESQ) measure [5], commonly accepted as the best objective speech quality measure on an 8-20 s scale. In particular, we modify PESQ leading to what we term PESQa which works in real-time on an internet Low Bit-rate Coder (iLBC) [6] 30 ms frame scale. Moreover, we propose a PBC Sliding Window Optimization (SWO), in which the PBC can adapt to time-varying changes in its pool of available PUPP resources. The combined result is an adaptable real-time working PBC based on high performance perceptual quality estimation. Secondly, we design a two channel protection scheme, Premium Channel (Pch) and Ordinary Channel (Och), in a VoIP DS protection setup and we test the connection between optimal PUPP resource allocation and network performance for varying loads of VoIP traffic given a single network point of congestion due to shortage of link capacity (i.e., link congestion). In this paper, our experiments are thus targeted at

This work was partially supported by the National Science Foundation via awards CCF-0347229 and CNS-0519933.

homogeneous VoIP traffic to consider the effects within this class of traffic as a first step towards a broader perspective. As we will show, in this setup, fixed PUPP resource allocation provides the best result. Additionally, packet loss statistics gives a clear indication on how to optimize PUPP from a channel design point of view.

The remainder of this paper is organized as follows. In Section 2, we present our improved PBC, and in section 3 we demonstrate the potential of the PBC, and we test the connection between network performance and PUPP optimization. Finally, we discuss the results and their implications in section 4.

## 2. PERCEPTUALLY BASED CLASSIFIER

Perceptually Based Classifiers assign packets to one of a set of $M$ available Transmission Channels (Tch's), given the perceptual importance of a packet and a constraint on resources as in Figure 1. We base our PBC on a variation on Rate-Distortion optimization similar in spirit to the approach taken in [3]. That is, let $\mathbf{a} = \{a_1, a_2, ..., a_N\}$ denote the channel assignment for each of $N$ packets in which $a_n$ is the assignment of the $n^{th}$ packet to one of $M$ Tch's (effected by either an amount of FEC protection or a DS priority level). Then the global optimal combination ($\mathbf{a_{opt}}$) is expressed mathematically by,

$$\mathbf{a_{opt}} = \underset{\mathbf{a}}{\operatorname{argmax}}\; Q(\mathbf{a});\ \text{for}\ \sum_{n=1}^{N} R_{a_n} \leq R_C, \tag{1}$$

where we choose to maximize expected quality instead of minimizing distortion. In Eq. (1), $Q(\mathbf{a})$ is the total *expected* quality at the *decoder* of the speech contained in the $N$ packets under a protection policy described by $\mathbf{a} = \{a_1, a_2, ..., a_N\}$. That is, under the given protection policy, one can determine various packet loss patterns and the resulting decoded signals and measure the quality of such decoded signals by comparing them to an original coded signal under a lossless transmission. In addition, $R_{a_n}$ is the cost given the choice of Tch, and $R_C$ is a constraint on overall cost. Therefore, Eq. (1) can be understood as determining the protection assignment $\mathbf{a} = \{a_1, a_2, ..., a_N\}$ that maximizes expected decoder quality subject to network-imposed resource constraints.

In [3], the paper did not examine the performance for time-varying amounts of allocated resources (i.e., changes on $R_C$). Therefore, we suggest a novel Sliding Window Optimization (SWO) approach whereby the PBC gains the flexibility to smoothly adapt to changes in the amount of available resources (i.e., $R_C$) for PUPP. Furthermore, to improve distortion/quality measurements, we adopt PESQ [5], generally accepted to have very high correlation with subjective speech quality assessment, and modify it to work as a real-time quality measure for evaluation of packet perceptual importance. We refer to this altered version as PESQa. First, we describe the Sliding Window Optimization.

### *2.0.1. Sliding Window Optimization*

SWO is an adaptive real-time operating approximation to global optimal PUPP, as described by Eq. (1). Therefore, initially, we describe how the global problem can be solved. Under a set of relevant assumptions this global problem translates to a global threshold applied to packet perceptual importance scores, see Figure 2. First we assume that quality is independent and additive across packets, which is reasonable for e.g. iLBC. Thus, $Q_L(n)$ denotes the packet playout quality of the $n^{th}$ assuming a packet loss. That is, $Q_L$ is the quality of the packet loss concealment measured relative to the playout signal following a reception (best achievable quality). In addition, $Q_R$ is the quality assuming a packet reception (a constant maximum quality level). Note, this assumption is justified by our choice of a frame-independently decodable coder, like the iLBC, though still quite coarse. Secondly, we restrict ourselves to two Transmission Channels, a Premium Channel (Pch), and an Ordinary Channel (Och), with the Pch providing better protection at a higher cost. In particular, we assume that the Tch's can be defined by fixed packet loss probabilities ($P_L(\text{Pch})$ for the Premium Channel, and $P_L(\text{Och})$ for the Ordinary Channel) and utilization costs ($R_{\text{Pch}}$ and $R_{\text{Och}}$). Note, for a given cost constraint, we assume both $P_L(\text{Pch})$ and $P_L(\text{Och})$ to be constant over time, though in reality these will vary slowly. Given these assumption, we first recognize that we may translate the original cost constraint into a target rate ($T_R$), where $T_R$ is the ratio of premium (ordinary to premium upgrades allowed by the original cost constrain) over total packets $N$. Hence we may express Eq. 1 as follows:

$$\mathbf{a_{opt}} = \underset{\mathbf{a}:\sum_{n:a_n=\text{Pch}} 1 = T_R \cdot N}{\operatorname{argmax}}\; Q(\mathbf{a}), \tag{2}$$

Secondly, we may express overall expected quality $Q(a)$ as the expected quality on a per packet basis given by the following equation:

$$Q(\mathbf{a}) = \sum_{n=1}^{N} \left( P_L(a_n) \cdot Q_L(n) + (1 - P_L(a_n)) \cdot Q_R \right), \tag{3}$$

in which $a_n$ can be equal to either Pch or Och. Finally, by substituting Eq. 3 into Eq. 2 and splitting terms according to the Pch and the Och, we get:

$$\begin{aligned}\mathbf{a_{opt}} = \\ \underset{\mathbf{a}:\sum_{n:a_n=\text{Pch}} 1 = N \cdot T_R}{\operatorname{argmax}} \Bigg\{ & P_L(\text{Pch}) \cdot \sum_{n:a_n=\text{Pch}} Q_L(n) + P_L(\text{Och}) \cdot \sum_{n:a_n=\text{Och}} Q_L(n) + \\ & Q_R \cdot (1 - P_L(\text{Pch})) \cdot \sum_{n:a_n=\text{Pch}} 1 + Q_R \cdot (1 - P_L(\text{Och})) \cdot \sum_{n:a_n=\text{Och}} 1 \Bigg\}\end{aligned} \tag{4}$$

Here the last two terms are constant over $\mathbf{a}$ for a given $T_R$ and consequently do not affect our optimization of $\mathbf{a}$. The first two terms tell us that we should assign packets such that the ones with the lowest $Q_L(n)$ (i.e. lowest quality assuming a loss, therefore perceptually important) are sent on the Pch, which has the lowest weight ($P_L(\text{Pch})$) in the expression. Moreover, the packets with the highest $Q_L(n)$ (best quality in case of a loss) are sent on the Och and the ratio is determined by $T_R$. In other words, given $T_R$, we can determine a global threshold in the $Q_L$ (perceptual importance) domain whereby we optimize PUPP on a global scale. This is illustrated on Figure 2.

Given the global solution, SWO follows from Figure 2. Now, instead of a global threshold, we estimate the global threshold within a Sliding Window (SW) of size $w$. Hence, for each new packet $n$ we estimate $Q_L(n)$, update the SW ($Q_L(n-w+1)$ to $Q_L(n)$), find the optimal threshold within the SW, and assign packet $n$ according to this threshold. Hereby, we achieve flexibility, approximating global optimization, with better protection for the perceptually important packets.

### *2.0.2. PESQa: Approximates PESQ*

Perceptual importance ($Q_L(n)$) is measured by PESQa that inherits most of its functionality from PESQ. That is, Figure 3 illustrates

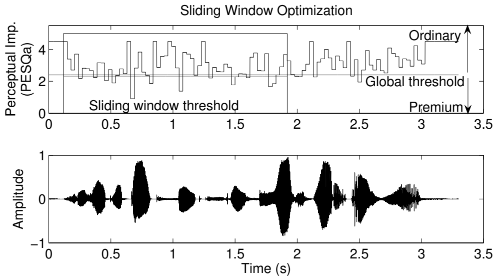

**Fig. 2**. Sliding Window Optimization (SWO). Global threshold approximated by sliding window threshold. In this example, the target is to assign the 20% packets with the lowest PESQa-Mos quality scores to the premium channel.

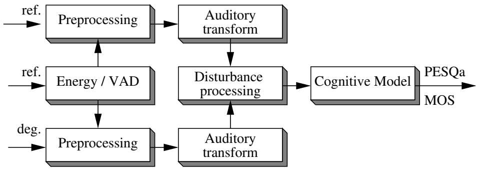

**Fig. 3**. PESQa block diagram.

the main functionalities with preprocessing, auditory transformation, disturbance processing, and cognitive model processing basically coming from PESQ [5]. With this setup, PESQa measures speech quality on a scale from -0.5 to 4.5 PESQa MOS, and, in general, the estimate is based on a comparison between a degraded and a reference speech sample, i.e. loss and lossless transmission in this paper.

Our PESQa is designed to operate as a real-time speech quality measure on the scale of iLBC 30 ms speech frames. This constitutes its main difference from PESQ, which works on a 8-20 s scale. This is possible by enabling PESQa to work on short speech samples, down to a few frames, and by removing unnecessary functionality, given its application. In particular, working on short speech samples creates a problem during preprocessing, where speech samples in PESQ are scaled to a target Average Power Level (APL) based on sample APL. For PBC operation, short speech samples leads to adversely diverged scaling across speech frames. Hence, for PESQa, we estimate long turn APL by first order low-pass filtering of short sample APL's. Relying on PESQ terminology [5], in PESQa the following PESQ functions have been omitted: time alignment, long term aggregation, transfer function equalization, and Intermediate Reference System (IRS) filtering. These changes are possible because: PESQa works on short time aligned speech samples, speech samples do not experience system filtering, and it is our conviction that IRS has lost its purpose in a VoIP framework, respectively. In addition to these changes, PESQa and iLBC frame sizes has been aligned such that PESQa splits input signals in 50% overlapping frames of 240 samples with 16 samples zero padding as the basic unit of comparison. PESQ relies on 256 sample frames. Also, we add a Voice Activity Detector (VAD) to PESQa. During PBC operation, VAD detected silence frames are per default given a maximum PESQa score.

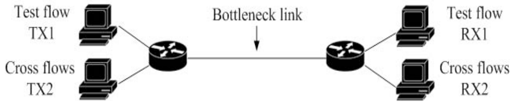

**Fig. 4**. Network topology for link congestion test.

## 3. EXPERIMENT

In this section, we study the connection between optimized PUPP resource allocation and network performance. To this end, we use the following setup. Figure 4 illustrates our choice of network topology (unless otherwise stated) in which we assume a single primary congestion point, a bottleneck. In particular, this bottleneck is caused by insufficient link capacity, given the load. The network exclusively carries VoIP traffic characterized by Poisson distributed inter arrival times at the bottleneck. Each packet contains an iLBC 30 ms encoded frame encapsulated by a RTP/UDP/IPv4 header and frames are generated based on sentences from the TIMIT database with an increased ratio of active speech (0.96, on average according to ITU-T P.56 measurements). According to our PBC, packets are sent on either a Pch or an Och implemented according to the IETF DS architecture. That is, assuming that all traffic share a common physical buffer, at the bottleneck, premium traffic is given better conditions by restricting the use of this buffer for ordinary traffic. In particular, ordinary traffic is dropped both if the physical buffer is full or according to a virtual Random Early Detection (RED) queue. Here, the virtual RED queue is practically being set up as a drop tail queue. Obviously, virtual queue size influences the performance of our protection scheme with optimal queue size depending on network congestion and premium rate. Consequently, every integer virtual queue setup has been considered such that all results presented are optimal in this respect. Physical buffer size is set to ten packets and all this is set up in the Network Simulator-2 (NS2) with one test flow and 9 cross flows, all allowed the same share of premium traffic. Speech quality is measured objectively with true PESQ, without IRS, averaged over sentence pairs and objective results are supported by a subjective listening test. We now present the setup of our PBC, including initial tests. Then, we present our results on optimized resource allocation supported by a subjective listening test.

### *3.0.3. Perceptually Based Classifier*

For experimental purpose, we use the following PBC setup. For each new packet, we compute the PESQa score comparing speech samples including the previous three packets where we assume either the loss or reception of the new packet, respectively. As part of this PESQa computation, the VAD mark the frame if sample APL is below 800 or below -33 dB compared to the previous estimated long turn APL (These values are chosen empirically for the TIMIT database speech files). Given the new PESQa score, we update the Sliding Window (SW) which covers 180 PESQa scores. Note, the SW is initialized with PESQa scores based on the current talker. Next, the SW threshold is adjusted (up or down) in steps of 0.001 PESQa-MOS until the SW Pch rate just passes the Pch target rate. However, if the current threshold gives a SW rate within 1% point of the target rate, this threshold is kept. Finally, based on this threshold, we assign the new packet to the Pch if its PESQa score is below the threshold and to the Och if it is above. Packets with a maximum PESQa score are per default assigned to the Och.

For this setup, PBC execution time is estimated to be below 15 ms per 30 ms frame, on average, on a Pentium 2.2 GHz proces-

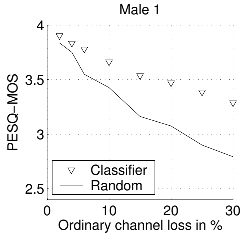

**Fig. 5**. PBC versus Random classification. Mean PESQ MOS scores as a function of ordinary channel loss rate.

sor. Further, due to SWO operation, the PBC do not exactly match a given Pch target rate. That is, for target rates up to 80%, the PBC premium assignment rate match target rates within ±2.5% point on a 3 s time scale. Also, the PBC has a maximum premium assignment rate of approximately 80% given that approximately 20% of all packets in our speech material receive a maximum PESQa score (per default assigned to the Och). However, most importantly, quality wise, we have tested our PBC against uniform random assignment and the results are illustrated in Figure 5. For this initial test of the PBC, we assume a lossless premium channel and losses on the Och are distributed according to a Gilbert-Elliot model with states received or lost, a mean burst length of 2, and varying loss rate, not in NS2. 44% of the packets are assigned to the Pch and PESQ is averaged over 2 sentence pairs repeated 40 times for both a male and a female speaker. We see that the PBC consistently improves speech quality, particularly for increasing ordinary channel loss rate.

### *3.0.4. Optimized resource allocation*

Figure 6 shows how speech quality varies as a function of the link capacity that varies between 200-270 $\mathrm{kb/s}$ (constant load). In addition, in each figure the premium channel assignment rate is also varied for target rates between 0% and 100% in steps of 10%. Note that 0% and 100% are equal as both represent EPP and 90% equals max assignment rate. The true PESQ scores are averaged over 2 Male and 2 Female speakers with 5 sentence pairs each, with the experiments repeated ten times. In general, PUPP gives better quality than EPP with approximately 70% PUPP leading to maximum quality consistently over all link capacities.

The PUPP improved performance comes from better control with packet losses. For example, consider the 200 $\mathrm{kb/s}$ scenario. Of the 70% of the packets sent to a premium channel, only 2% of these premium packets are lost while 72% of the ordinary packets are lost (in the given RED scheme) giving a total loss of 23%. In contrast, EPP (no protection) looses 17% of the previously marked premium packets and 17% of the previously marked ordinary packets giving a total loss of 17%. Clearly, the improved performance in PESQ comes from protecting the perceptually important premium packets, given that PUPP has a higher total loss rate.

### *3.0.5. Subjective listening-test*

The Objective PESQ-MOS results are supported by an informal Degradation Mean Opinion Score (DMOS)listening test. For this test, speech sentences were collected from the link congestion setup, at maximum congestion, applying either EPP or 70% PUPP and presented to 10 naive listeners.

Results are listed in Table 1, considering a 99% confidence interval. The results reinforce the earlier objective performance improvement due to 70% Premium Channel PUPP in a DS environment.

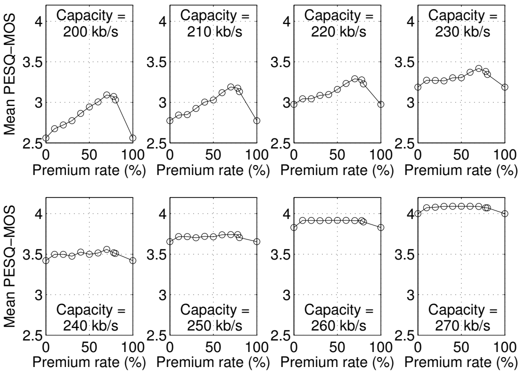

**Fig. 6**. PUPP PESQ-MOS scores as a function of premium rate and link capacity. Constant load of 1 test flow and 9 cross flows.

**Table 1**. Mean DMOS scores with 99% confidence intervals.

| Method | Mean MOS |
|---|---|
| EPP | $2.51 \pm 0.15$ |
| PUPP | $3.82 \pm 0.15$ |

## 4. DISCUSSION

In this paper, we have studied the effect of Perceptually Unequal Packet Protection (PUPP) on both voice quality and overall network performance. To this end, we have developed a novel real-time working perceptually based classifier (PBC) designed to work with a Premium channel (Pch)/Ordinary channel (Och) protection scheme. In particular, the classifier implements PESQa, our modified version of the objective speech quality measure PESQ, and SWO, a sliding window approximation to global optimization. Given this PBC and Differentiated Service based Pch and Och, results show that PUPP performs optimally at a 70% premium rate irrespective of network performance (in our setup), considering congestion due to a bottleneck link. However, packet loss statistics indicate that improved quality comes from a tradeoff between overall packet loss and ensuring that only the perceptually least important packets are lost. This indicates that we might optimize PUPP given a Pch and Och such that no premium packets are lost and overall loss rate is in direct ratio to bottleneck overload without protection. To this end, a topic of our current research is to investigate the effect of encoder based perceptually discarding of speech frames. Also, we are going to improve PBC performance by improving the distortion measure and by considering speech quality inter frame dependencies.

## 5. REFERENCES

[1] J. C. DeMartin, “Source-driven packet marking for speech transmission over differentiated-services networks,” *Proc. IEEE ICASSP*, vol. 2, pp. 753–756, 2001.

[2] S. Præstholm, S. S. Jensen, S. V. Andersen, and M. N. Murthi, “On packet loss concealment artifacts and their implications for packet labeling in voice over ip,” *Proc. IEEE ICME*, vol. 3, pp. 1667–1670, 2004.

[3] M. Chen and M. N. Murthi, “Optimized unequal error protection for vice over ip,” *Proc. IEEE ICASSP*, vol. 5, pp. V – 865–8, 2004.

[4] H. Sanneck, N. Le, A. Wolish, and G. Carle, “Intra-flow loss recovery and control for voip,” *Proc. ACM Multimedia*, 2001.

[5] International Telecommunicatio Union (ITU-T), *P.862, Perceptual Evaluation of Speech Quality (PESQ)*, 2001.

[6] S. V. Andersen et. al., *RFC 3951, Internet Low Bit Rate Codec (iLBC)*, IETF, 2004.