Development of an objective speech quality measurement model for the AMR codec

Stefan Wänstedt, Jonas Pettersson, Tan Xiangchun, Gunnar Heikkilä
AWARE, Ericsson Research
Ericsson Erisoft AB, Box 920, 971 28 Luleå, Sweden
Fax: +46 920 202099
{stefan.wanstedt, jonas.pettersson, xiangchun.tan, gunnar.heikkila}@epl.ericsson.se

Summary

This paper describes the development of a non-intrusive measure of speech quality for GSM AMR, based on bit error (BER) and frame error (FER) distributions together with handover events. Input to such a model is extensive listening tests on speech from AMR GSM networks to identify what type of error situations that cause poor speech quality. By using suitable mathematical/ statistical tools on the results from the listening tests and the full information about the errors and their distributions, a model of speech quality could be established. The cross-validated correlation coefficient of the model and subjective speech quality is 0.94.

Introduction

A cellular operator struggles with a wide variety of tasks concerning the optimization of the overall network quality. The activities comprise, among others, daily supervision, net tuning and locating low performance areas. As the telecom markets develop and competition increases, the importance of being able to provide a high service level towards the customers also increases [1]. The subjective quality perceived by the user is probably the most important Quality of Service (QoS) parameter when choosing an operator, especially when accessibility and services provided are comparable. Until now and probably for some time to come, speech has been and will continue to be the most important service in cellular systems. Hence, operators often use speech quality as a means for optimization of system performance. Since the standardized measures available in the system today, i.e. RxQual, cannot measure this critical parameter the operators require other means to acquire logs of speech quality.

Speech quality in current mobile systems is limited by the performance of the speech coder algorithms, signal processing elements such as echo-cancellers, design of terminal as well as by distortions caused by channel errors under varying radio conditions. In a well-controlled environment with good coverage, the speech coder itself may limit the quality, while in most environments the speech quality depends on the radio link quality. In most 2G systems the radio interface is by far the most important source of deterioration of speech. Furthermore, the radio interface is a part of the network controlled by the operator and represents the section where a majority of the optimization activities are performed.

In this paper we describe the development of a non-intrusive measure of speech quality for the Adaptive Multi-Rate (AMR) Codec. The model is based on the relationship between radio interface parameters and subjective speech quality. The model is especially suitable for tuning and optimization of air-interface related speech quality issues.

The AMR codec

The ongoing quest for improved speech quality, together with the need for improved efficiency of cellular networks, has resulted in a new speech codec. Since there appeared to be a limit to what fixed rate solutions, such as FR and EFR, could provide an adaptive solution was selected [1], called Adaptive Multi-Rate (AMR) [2]. The AMR codec comprises 14 different (6 half rate and 8 full rate) modes. The full rate modes have net bit-rates (kbit/s): 4.75, 5.15, 5.9, 6.7, 7.4, 7.95, 10.2 and 12.2, respectively. The six lowest of these bit-rates are supported for both full-and half-rate channels [2]. Even though AMR was primarily developed for GSM, the speech codec has been adopted for many of the coming cellular systems.

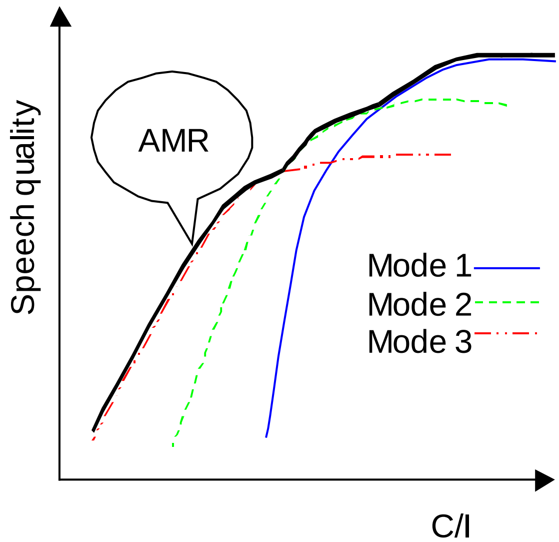

Figure 1. Principle of AMR mode changes to maintain speech quality when C/I decreases.

If the radio link quality deteriorates during a conversation the AMR codec changes to a mode with lower bit rate, but better error protection, Figure 1. In this way the connection can be maintained with acceptable speech quality even when the radio environment is severely disturbed. Mode changes are possible every second speech frame or every 40 ms.

Speech quality measurements

Speech quality in cellular systems can be measured intrusively or non-intrusively. The former commonly involves sending speech through the network and comparing the resulting speech with the original. The latter, the non-intrusive measurement in this case involves measuring some quality of the air interface that is highly correlated with speech quality.

Intrusive measurements

The intrusive measure is truly end-to-end, including radio network, core network, switches and user equipment, Figure 2. Examples of methods used for intrusive objective measurements are PSQM [3] and PESQ [4].

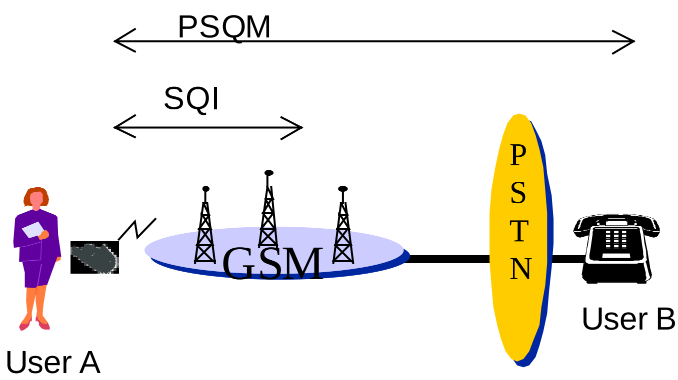

Figure 2. Range of E2E (e.g. PSQM) vs. air interface measurements (SQI).

Drawbacks with the intrusive methods are that only a single connection can be measured at a time and that the actual measurement adds load to the network. The measured quality applies only to the call used to transmit the speech sequence. Furthermore, the measurement chain includes parts not directly related to the radio network making optimization of the cellular network difficult, Figure 2.

Non-intrusive measurements

Traditionally in GSM systems, the receive quality (RxQual) has been used to measure the speech quality in networks. RxQual is part of the GSM standard measurement report sent from the mobile to the base stations and is therefore readily available. This availability makes it possible to supervise all calls in the network using RxQual. Unfortunately, RxQual does not do a very good job in evolved GSM systems, e.g. systems with frequency hopping. Future improvements to GSM systems, such as the Adaptive Multi-Rate (AMR) speech codec, will include changes of the efficiency of error correction mechanisms, resulting in changed correlation between RxQual and the true network quality.

In an attempt to improve speech quality measurements the Speech Quality Index (SQI) was developed [3]. The SQI, based on bit error (BER) and frame error (FER) distributions together with handover events and use of DTX, Figure 3, predicts the instant speech quality in a phone call/radio-link in real-time. SQI has been implemented in drive testing tools as well as in the system performing up-link measurements of speech quality. SQI is calculated over 125 speech frames or 2.5 s to obtain a high resolution in time, e.g. along a drive test route. The correlation between SQI and subjective speech quality is about 93% for 2.5 s samples, but is as high as 98 % for 10 s samples [3]. Hence, the performance of this non-intrusive method is comparable to e.g. PSQM, in the context it is designed for.

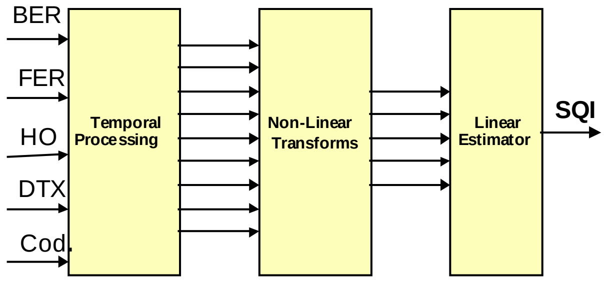

Figure 3 Principle of SQI, input parameters and basic processing steps.

Developing a speech quality model for AMR

The adaptive nature of AMR makes speech quality measurements based on the radio environment more difficult. Due to mode changes the relations between radio link parameters included in SQI also change. Hence it is necessary to quantify this information to adequately model speech quality.

SQI models in general

The development of the SQI models requires extensive listening tests on speech from the network, in this case GSM, to identify what type of error situations cause poor speech quality [3]. The results from the listening tests and the full information about the errors and their distributions are used as input to the model, Figure 4. To make sure that subjective grading is consistent and repeatable, all samples are graded with respect to MNRU [3], that is, the dBQ scale is used.

The radio quality data are analyzed to determine what parameters are important, their distributions as well as their interrelations. If necessary the parameters are transformed. For example, BER is linearly related to speech quality and can be used as is, while FER behaves non-linearly and requires some transformation. For example, the square root of FER correlates well with speech quality. Another important measure based on FER is the longest consecutive sequence of frame losses (LFE). Both FER and BER are required to provide speech quality estimates with high resolution over a wide C/I range. FER works well for low C/I, while BER dominates on the other end of the C/I scale.

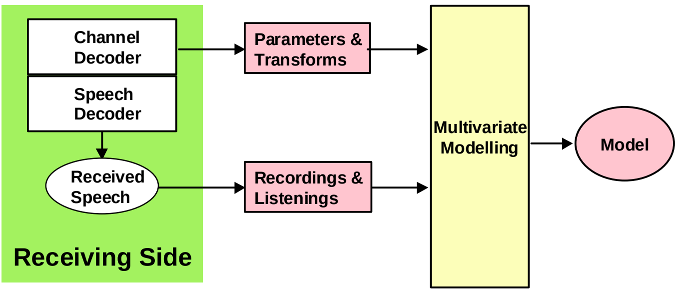

Figure 4. Schematic of receiver side of modeling sequence.

Finally, a multivariate statistical model of speech quality is developed. Further transformations of variables may be necessary to produce the best model possible. The general form of SQI is:

$$A * BER + B * FER^x + C*LFE + Const.$$

The AMR model

Since AMR has not been launched yet, all tested speech sequences were simulated. Eight different speech sequences were used with 4 different speakers each reading 2 sequences. Four different channel fading models, with different types of frequency hopping, were used to add air interface errors. The C/I range of the different channels was –6 to 19 dB.

Due to the very large set of simulated speech sequences, totaling almost 26000 samples, grading all speech sequences was not possible. Hence a representative selection of samples had to be made to ensure that no part of the data space was unevenly represented since this may result in a bias. In this particular case, the majority of data were reasonably high quality and noise-free. Consequently, a large part of the good quality data had to be removed from the data set. The selection process is quite easily understood by an illustration in one dimension, Figure 5, where the distribution is flattened by cropping the histogram above a certain level.

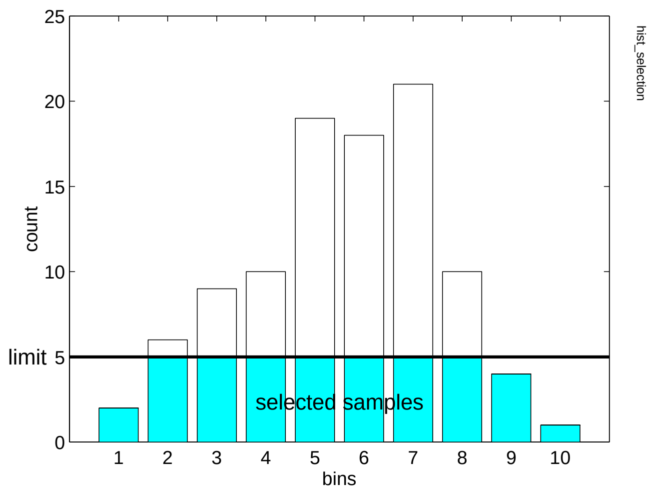

Figure 5. Principle of sample selection in one dimension. The selection process evens the distribution by cutting the top of the histogram

Analyzing the radio link parameters gives that the relation between speech quality and BER for the AMR codec changes with mode, while no such mode dependency can be seen for the other radio link parameters. The simple solution, to omit BER from the model, would result in a model with a much lower resolution on good channels. In other words, this is not possible. Consequently, a model per mode would be required unless this problem could be solved.

Based on the relation depicted in Figure 6 a normalized BER (NBE) was calculated, Figure 7. The NBE basically transforms the data clusters for each mode so that the slopes coincide. Assuming that this is a linear transform NBE can be defined as:

NBE = *a* + *b*BER*, for each speech frame.

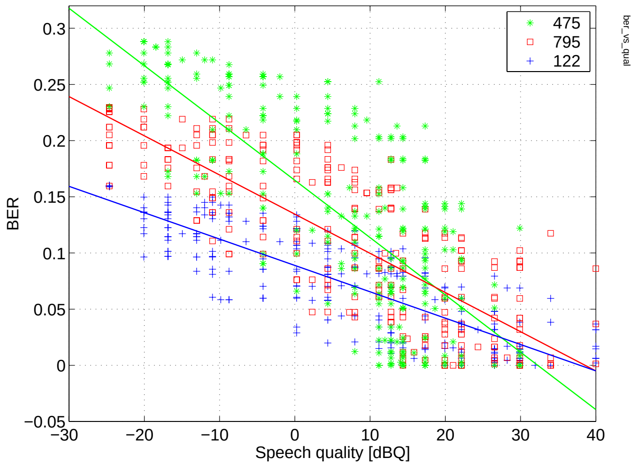

Figure 6. Relation between speech quality, mode and BER.

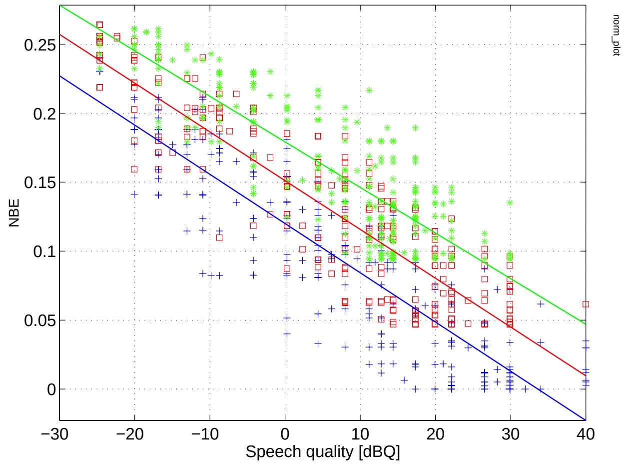

Figure 7. Normalized bit-errors before optimization.

Initially this calculation was performed manually using basic algebra. These preliminary values were then tuned by optimizing the model performance, by changing the values using non-linear least-squares data-fitting. From this normalized NBE a single model capable of estimating SQI for all 14 modes could be developed.

The model is based on the same radio link parameters as previously, except for the normalized BER.

Hence, the relation between radio link parameters in final model can be described as:

SQI = A**NBE* + B**FER*^{x} + C**LFE* + *Const.*

The values *a* and *b*, used to calculate *NBE*, are functions of the AMR mode and are defined per speech frame, while SQI is calculated every 2.5 s (125 frames).

Model performance

A comparison with subjective quality and another speech quality measure shows that the relation between speech quality and radio link parameters is fairly robust.

Compared to subjective grades

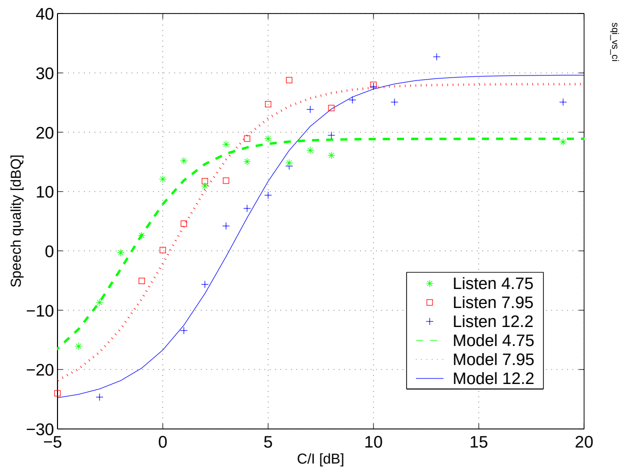

Figure 8. Examples of AMR modes, model and average subjective grades. Speech and radio-link data are simulated.

Subjective scores are “true” speech quality since they represent the quality really perceived by the user. Figure 8 shows examples of SQI for 3 modes calculated with the model and the respective subjective scores. Each grade represents an average of several speakers and listeners. The cross-validated correlation coefficient of 0.94 shows that the model describes speech quality well. Figure 9 shows the performance of the SQI model.

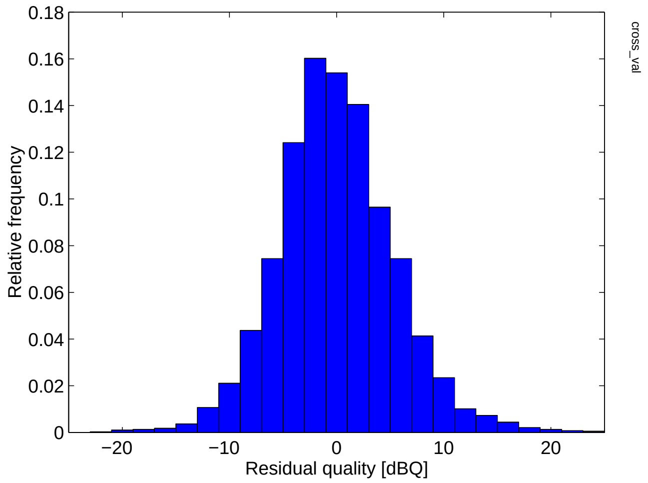

Figure 9. SQI model compared to speech quality.

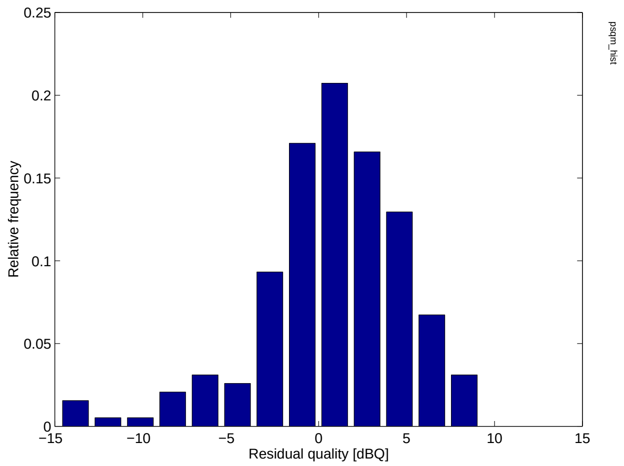

Figure 10. PSQM scores compared to speech quality.

Compared to PSQM

PSQM is a common intrusive objective speech quality measure. Comparing PSQM scores to the speech quality of the test set gives an idea of how the non-intrusive measure compares with an intrusive measure. Figure 10 shows a similar plot as Figure 9 but based on PSQM scores of the data set. The correlation of this particular data set is –0.93.

Conclusions

The speech quality index described in this paper provides excellent estimates of speech quality for the AMR codec planned to be used in coming releases of GSM. Although AMR adds difficulty by allowing mode changes as often as very second speech frame, SQI will be able to calculate the speech quality for 2.5 s sequences. The performance of the model is good compared to subjective grades, with a correlation of 0.94 for the tested data set. As a comparison PSQM was run on the data set as well. The correlation of PSQM and subjective speech quality for the tested data set was –0.93.

References

[1] Minde, T.B., Bruhn, S., Ekudden, E., and Hermansson, H. “Requirements on Speech Codecs Imposed by Speech Service Solutions in Cellular Systems.” IEEE Workshop on Speech Coding, September 1997.

[2] Uvliden, A., Bruhn, S., and Hagen, R. “Adaptive Multi-Rate: A Speech Codec Adapted to Cellular Radio Network Quality.” ASILOMAR 1998.

[3] ITU-T Recommendation P. 861. “Objective quality measurement of telephone-band (300-3400 Hz) speech codecs. February 1998.

[4] ITU-T Recommendation P.862. “Perceptual Evaluation of Speech Quality (PESQ), an objective method for end-to-end speech quality assessment of narrowband telephone networks and speech codecs”. May 2000.

[5] Karlsson, A., Heikkilä, G., Minde, T. B., Nordlund, M., and Timus, B. “Radio Link Parameter Based Speech Quality Index – SQI”. IEEE Workshop on Speech Coding, June 1999.

[6] ITU-T Recommendation P. 810, “Modulated Noise Reference Unit, MNRU”.