Adaptive Thresholds for AMR Codec Mode Selection

Tomas Lundberg, Peter de Bruin, Stefan Bruhn, Stefan Häkansson, and Stephen Craig
Ericsson Research
Ericsson AB
Sweden
Email: {tomas.lundberg, peter.de.bruin, stefan.bruhn, stefan.lk.hakansson, stephen.craig}@ericsson.com

Abstract—The speech codecs from the Adaptive Multi-Rate (AMR) codec family enable provisioning of excellent speech quality, at the same time providing a way forward towards state-of-the-art, spectrally efficient, high capacity cellular networks. One straightforward way to characterize the benefit of AMR speech codecs is that the robustness to interference and noise in radio networks is increased and that this advantage over other, non-adaptive, speech codecs can be capitalized on in several different ways, e.g., by enhancing speech quality or improving system capacity. In this paper, improved mode adaptation, where codec mode switching thresholds are adaptive to radio conditions, is discussed. Example simulations show that an adaptive thresholds algorithm applied to GSM can significantly improve objective speech quality. Corresponding improvements were also found in informal listening tests.

I. INTRODUCTION

Adaptive Multi-Rate (AMR) codecs are standardized by 3GPP for GSM [1], [2], the world’s most widespread cellular technology, as well as for WCDMA. In GSM the codec mode adaptation is based on estimations of the radio link quality, while in WCDMA the AMR adaptation concept works completely differently. Quality of Service control in WCDMA basically corresponding to AMR codec mode adaptation in GSM is done by means of fast power control, whereas codec mode adaptation may be used as a tool for capacity control. In this paper we will exclusively deal with AMR in GSM.

Two variants of AMR exist, narrowband AMR and wideband AMR (AMR-WB). Narrowband AMR consists of eight codec modes with different source bit rates, from 12.2 kbps down to 4.75 kbps. It provides the traditional audio bandwidth of PSTN telephony of about 100–3500 Hz. AMR-WB contains nine different codec modes with source bit rates from 6.6 kbps up to 23.85 kbps, and with an audio bandwidth of 50–7000 Hz. The increased bandwidth improves the intelligibility and naturalness of speech significantly at the same time as the quality for music and mixed content material is improved.

Common to both AMR variants is that a number of codec modes are collected into a pre-defined Active Codec Set (ACS), which usually is fixed during a call. The level of channel coding is adjusted depending on the source bit rate so that the total (gross) bit rate is constant. Consequently, the lower the source bit rate becomes, the more robust the codec mode is against bit errors, since a larger portion of the bit rate is used for channel error protection.

The gains from AMR are two-fold: When radio quality is sufficiently good, a higher codec mode can be used, thereby improving the perceived speech quality. On the other hand, since lower AMR codecs modes are very robust, radio network planning can be made to allow an increased interference level in the system, thereby increasing system capacity significantly, comparing, e.g., narrowband AMR to the common Enhanced Full Rate (EFR) codec in GSM. Thus, introducing AMR may improve speech quality, increase capacity or yield a combined effect of both quality and capacity improvement.

This paper proposes Adaptive Thresholds as an improvement to current AMR codec mode adaptation. Section II describes mode adaptation in more detail and motivates the need for adaptive thresholds. An algorithm is exemplified in section III, and simulations in section IV analyze its impact on speech quality. Finally, conclusions are outlined in section V.

II. CODEC MODE ADAPTATION

A. General description

For codec mode adaptation, the receiving side performs radio link quality measurements of the incoming channel yielding a quality indicator (QI), which is defined as an equivalent carrier-to-interference (C/I) ratio [3]. The QI is then compared to a set of pre-defined thresholds to decide which codec mode to use. The thresholds are normally fixed during a call, but the system can initiate a change to the thresholds.

To obtain the best possible speech quality it is important to properly select the thresholds for codec mode adaptation. This implies that it is necessary to obtain a QI that correctly reflects the speech quality for any given radio condition, frequency hopping scheme, and network configuration. Obviously, this procedure may be quite complicated. Furthermore, conditions vary over time. There may also be performance variations between different receiver units, both regarding actual performance and QI estimation. This means that it is likely that even well-selected adaptation thresholds will not be optimal at all times.

Fixed thresholds can be sub-optimal for the current conditions by being either too high or too low. In the case where
the thresholds are too high, a switch from a less robust codec mode to a more robust mode will be initiated earlier than necessitated by the radio conditions. This will cause a slight degradation of the speech quality due to the lower intrinsic speech quality of the more robust codec mode. A more serious problem arises when the thresholds are too low, causing the switch from the less robust mode to occur too late. This may significantly increase both the frame and bit error rates after channel decoding and in turn cause a severe degradation of the speech quality. However, since both cases lead to speech quality reduction they should both be avoided.

B. Adaptive thresholds for codec mode adaptation

Our proposed solution to the above-mentioned problems is to use thresholds that are adaptive to the current radio conditions. An algorithm to adapt thresholds could be applied either on the terminal side or on the network side, working on the uplink and/or the downlink. Hence, considering downlink threshold adaptation, either the receiving terminal could modify the thresholds directly, or the network could initiate the threshold adaptation in the terminal based on the measurement reports that it receives. The current standard allows the application of a “normalization factor” to normalize the QI for different receiver performances. So even though it could be argued that terminals are not allowed to modify their own codec mode switching thresholds, a direct modification and adaptation of thresholds in the proposed manner is considered to be standard compliant.

There are in theory several advantages of having the algorithm working on the receiving side, i.e., in the network and working on the uplink or in the terminal and working on the downlink, since in that case more information about the quality of the link can be made available to the algorithm. It will lead to a faster and more accurate adaptation of the thresholds. However, in practice it is often desirable for operators to align the QI estimation and hence adaptation performance of the various Mobile Stations (MSs) from different vendors, which are likely to vary. This implies that for downlink threshold adaptation an algorithm where the network initiates the threshold adaptation in the MS is required.

For such a network-based algorithm working on the downlink the new set of thresholds could be sent to the MS using a RATSCCH message (Robust AMR Traffic Synchronized Control Channel). A RATSCCH message steals one speech frame in AMR Full Rate and two speech frames in AMR Half Rate, effectively causing one or two erased frames. Consequently, the algorithm should not be allowed to update the thresholds too often. Simulations show, however, that this is not a problem in realistic scenarios.

III. DESCRIPTION OF THE ALGORITHM

In this paper we propose a network-based threshold adaptation algorithm for the downlink. As this case is more constrained than the case where the threshold adaptation is done directly at the receiver, its performance will be a lower bound for direct adaptation methods where the adaptation is performed in the same node as that in which the radio link quality measurements are conducted.

The proposed algorithm estimates the speech quality on the receiving link (i.e. in the MS) and compares the estimate against given speech quality limits for each codec mode. A simplified block diagram of the algorithm can be seen in Fig. 1. The speech quality could be estimated from Frame Erasure Rate (FER), bit error rate measures, e.g., RxQual, or objective speech quality measures, e.g., SQI [4]. In this particular implementation we have elected to use FER reported by the MS in the Enhanced Measurement Report (EMR) [5]; consequently, a necessary requirement for the algorithm to work is that the MS supports EMR. EMR includes the number of correctly received frames during each measurement period of 480 ms, i.e., 24 speech frames, from which the FER can be derived.

The speech quality estimate for the receiving link is generally too noisy for direct use in the threshold adjustment decision, and instead a long-term average is calculated. If the estimated long-term speech quality for a given mode is outside its limits, either too good or too bad, it is likely that the associated threshold for switching to the appropriate adjacent codec mode is sub-optimal for the current radio conditions. The algorithm will then modify all codec mode switching thresholds accordingly. The reason for modifying all thresholds instead of only the threshold in question is mainly practical; it simplifies maintaining the thresholds in a consistent order, i.e. not overlapping [3].

It is, in fact, necessary to estimate two different long-term speech quality values, one for comparing with the upper speech quality limit, and one for comparing with the lower limit. The reason for this is that a high FER value in the lowest, most robust codec mode is obtained when the radio conditions are poor, regardless of the values of the thresholds. Therefore, FER values obtained when in the lowest codec mode are discarded when calculating the long-term speech quality used for comparison with the upper limit. The converse is true for the long-term speech quality used for comparison with the lower limit, i.e. regardless of the value of the thresholds, a low FER value in the highest, least robust codec mode is obtained when the radio conditions are good, and consequently FER values obtained when in the highest codec mode are discarded in this case.

The calculation of the threshold adjustment is based on the two long-term speech quality estimates and on the attainable

---

**Fig. 1.** Simplified block diagram of the adaptive thresholds algorithm. SQ denotes “speech quality”.
speech quality with the current ACS. Due to the averaging process used in obtaining the long-term speech quality estimates, several threshold adjustments will usually have to be made. Thus, the threshold adaptation speed is much lower than the codec mode adaptation. The threshold adaptation can be thought of as a fine-tuning of the normal codec mode adaptation: the normal codec mode adaptation takes care of the time-critical adaptation in response to rapid changes in the radio environment. Concurrently, the threshold adaptation monitors the normal codec mode adaptation, and compensates for long-term systematic errors.

IV. SIMULATION RESULTS

Simulations were performed in an AMR link simulation tool, using Typical Urban 3 km/h (TU3) as the fast fading profile, additional slow fading (shadowing), and with different frequency hopping scenarios. An ACS with the modes MR475, MR59, MR67, and MR102 was used to evaluate the performance gain, where a receiver with incorrect C/I estimation was simulated. The incorrect C/I estimation could for instance be due to unusual radio environments, broken antennas, a stationary MS in a non-frequency hopping environment, etc. Moreover, since the standard does not contain any detailed description on implementation but merely lists performance requirements, another cause for inconsistencies in the C/I estimation could be different, though standard compliant, receiver and C/I estimator implementations.

For simplicity and ease of interpretation, a fixed, systematic error was studied. However, it should be noted that the algorithm is not limited to deal with only this kind of error. The incorrect C/I estimations were modeled by adding a positive or negative offset to the C/I estimates before the codec mode selection. The offset was kept fixed during a given simulation run, and varied between different runs. A positive C/I offset (i.e. an overestimation of the C/I) corresponds to setting the thresholds too low, while a negative offset (i.e. an underestimation of the C/I) corresponds to setting the thresholds too high. As previously discussed, using too low thresholds is a more serious problem than using too high thresholds. Consequently, more emphasis was put on evaluating the former case.

The simulations were run for 22,000 speech frames (440 seconds) and the speech quality was evaluated using PESQ, a tool for calculating objective speech quality [6], as well as informal listening tests. FER statistics and codec mode usage were also registered and used in the evaluation.

A. Overestimated C/I

The speech quality evaluation results for TU3 with Ideal Frequency Hopping (IFH) are shown in Fig. 2, where “correct” means correct C/I estimation, “+2 dB” an overestimation of 2 dB, etc. As can be seen in the figure, when employing the adaptive thresholds algorithm, the speech quality is maintained at approximately the same level as without any offset, even at the higher C/I estimation error levels.

Fig. 3 shows a detailed plot of the speech quality for each segment of 8 seconds for the case of an overestimation of 5 dB. For each such segment the speech quality is given as the difference in speech quality relative to the corresponding segment in a simulation with the correct C/I estimation. The speech quality for the first segment (given by the first two bars from the left) is considerably lower compared to a simulation using correct C/I estimation, and this is caused by not selecting a sufficiently robust codec mode due to the overestimation of the C/I. At this initial stage the threshold adaptation is not yet in steady state. The first threshold adjustment occurred after approximately 5 seconds, and after that the speech quality, on average, improved considerably. The fluctuations in speech quality, and the seemingly strange phenomenon that the speech quality without adaptive thresholds can be higher than with adaptive thresholds for a few segments, have a simple explanation: when codec mode switching thresholds are determined, the optimal thresholds are those that are optimal over the long term, i.e. on the order of minutes. The eight second long segments used in Fig. 3 are so short that, due to random fluctuations in the radio environment, a codec mode other than the globally optimal codec mode can give better speech quality for just that particular segment. The aggregated speech quality, e.g. as shown in Fig. 2, is obtained by averaging the segments over time.

The corresponding view in the FER domain is shown in Fig. 4. Here the accumulated number of frame erasures is plotted against time, with and without adaptive thresholds, together with the location and size of the threshold adjustments. The first threshold adjustment, in which the thresholds were adjusted 3 dB upwards, came after only 4.8 seconds. A second adjustment of 2.5 dB occurred after a little less than two minutes. In this particular case, the algorithm slightly overcompensated the C/I overestimation of 5 dB, since the thresholds were adjusted by in total 5.5 dB. This was due to a deliberate design choice: it was deemed more important to make rather large steps in the threshold adjustments in order to speed up the adaptation than it was to avoid the risk of overcompensating the C/I estimation bias. Such an
Fig. 3. Speech quality for each 8 second segment for the case of a 5 dB overestimation. The speech quality is given as the difference relative to a simulation with correct C/I estimation. Only the first 300 seconds are shown.

Fig. 4. The accumulated number of frame erasures (FE) for an overestimation of 5 dB, with and without adaptive thresholds, compared to a simulation with correct C/I estimation. The arrows show the location and the size of the threshold adjustments.

approach also reduces the number of threshold adjustments that have to be made, which means less signaling, although at the potential cost of slightly suboptimal thresholds in some situations. It is unrealistic to expect the algorithm to always exactly compensate for an incorrect C/I estimation; however, depending on the algorithm parameter settings, it almost exclusively comes within ±1 dB.

B. Underestimated C/I

The results from simulations with an underestimation of the C/I are more subtle. This is to a large extent due to the fact that the difference in speech quality in this case almost exclusively comes from selecting a too robust codec mode, which gives a FER level that is too low. In other words, the difference in speech quality in this case is mainly due to the difference in intrinsic speech quality of the different codec modes. This is in contrast to the case of C/I overestimation where the differences in speech quality from not selecting sufficiently robust speech codec modes are much more pronounced. Also, in contrast to the overestimation case, the algorithm here must not be too aggressive in its threshold adjustment, since adjusting the thresholds downwards too much must be avoided at all costs.

Fig. 5 shows the speech quality averaged over the whole speech database for simulations with different levels of the C/I underestimation. Using adaptive thresholds improves the speech quality, although not as drastically as in the case of overestimating the C/I.

The explanation for the improvement in speech quality can be found by looking at the relative usage of the different codec modes in Fig. 6. Here it can be seen that with adaptive thresholds the usage of the highest mode approaches the level obtained with a correct C/I estimation, and that the usage of the lowest mode has decreased. Despite this clear improvement in codec mode usage, the use of adaptive thresholds does not quite enable the optimal codec mode distribution as obtained with a correct C/I estimation to be reached. This is why the average speech quality with adaptive thresholds in Fig. 5 does not reach the level obtained with a correct C/I estimation.

The higher relative usage of the MR475 codec mode with its lower intrinsic speech quality lowers the overall speech quality compared to the case with correct C/I estimation. However, as discussed previously, the adaptive threshold algorithm must be rather conservative in its adjustments when compensating for underestimations in C/I. The FER level after threshold adjustments should never be allowed to exceed the FER level obtained with correct C/I estimations since that would lead to significant reductions in speech quality.

C. Listening tests

To verify the promising simulated speech quality results, informal listening tests were also performed. For all simulated
V. CONCLUSIONS

Codec mode switching thresholds that are adaptive to radio conditions, referred to as Adaptive Thresholds, have been proposed as a means to further improve AMR mode adaptation. Simulations have shown that an adaptive thresholds algorithm applied to AMR in GSM can improve objective speech quality significantly. The improvements have also been confirmed in informal listening tests.

REFERENCES


Fig. 6. The relative codec mode usage for the simulation with a 5 dB underestimation, with and without adaptive thresholds, compared to the relative codec mode usage for the case with correct C/I estimation. In all three cases the statistics are collected starting at the speech frame corresponding to the last threshold adjustment in the simulation with adaptive thresholds.

cases, for overestimated as well as underestimated C/I measurements, the speech quality improvements were confirmed. Compared to the case without adaptive thresholds, the relative improvements could be readily perceived.