eAMR: Wideband Speech over Legacy Narrowband Networks

Stéphane Villette, Sen Li, Pravin Ramadas, Daniel J. Sinder

Qualcomm Technologies, Inc., 5775 Morehouse Drive, San Diego, CA 92121-1714, United States {svillett, senl, pramadas, dsinder}@qti.qualcomm.com

Abstract

2. BACKGROUND

This paper introduces eAMR (enhanced-AMR), a novel technique for delivering wideband speech over existing narrowband networks. Instead of using a completely new wideband speech coder which would require new infrastructure, as is the case e.g. for AMR-WB or EVS, eAMR is based on the existing AMR (narrowband) codec, which is already widely deployed. eAMR uses an efficient coding model to represent the high frequencies of the speech signal, and combines it with watermarking technology to hide this data within a normal narrowband AMR bitstream. As a result, eAMR is a wideband codec which is fully compatible with the existing AMR network infrastructure, and therefore can be deployed as a handset-only feature.

Index Terms— speech coding, watermarking, wideband, AMR, AMR-WB, EVS

1. Introduction

Until a few years ago, the quality of voice telecommunications has been limited by design choices made over 100 years ago, which resulted in an 8 kHz sampling rate being used, for a practical frequency range of 300-3400Hz. This so called Narrowband (NB) frequency range severely limited speech quality. Recently, the industry has started to move to "HD voice" and "Ultra HD voice", i.e. the use of wideband (WB) or super-wideband (SWB) coders, respectively, which use a sampling rate of 16kHz or 32kHz and correspond to a frequency range of 50-7000Hz or 50-14000Hz respectively [1] [2].

This move has been slow, due to the fact that a whole new infrastructure is needed to support these wideband coders, at substantial cost.

This paper presents Qualcomm Technologies' enhanced-AMR (eAMR) solution, which combines efficient high-band coding and watermarking techniques to provide WB speech on top of the existing AMR NB coder. eAMR-enabled handsets are able to communicate in WB, while being fully backwards compatible with the existing NB networks and devices. Therefore, eAMR allows WB to be deployed on commercial networks without any infrastructure changes. Ever since the beginnings of digital telephony, it has been accepted practice that voice could be acceptably represented at a sampling rate of 8 kHz, and that a frequency range of 300-3400Hz was sufficient for good quality speech. Subsequent systems, including digital speech coding systems, have simply been designed to be compatible with these older systems [1][5]. Current such state-of-the-art NB codecs include the EVRC and EVRC-B codecs for CDMA networks [3], and the AMR codec for GSM/UMTS networks, which offer good quality typically at 4 to 12 kb/s. Improvements in signal processing, DSP, and electroacoustics, have now made possible the use of a wider frequency range. It is now well known that frequencies

frequency range. It is now well known that frequencies above 3400 Hz are very useful for speech intelligibility, speech quality, and reducing listener fatigue. Certain sounds, for example 's' and 'f', are much easier to distinguish when high frequencies are included.

As a result, there is currently a lot of interest in the speech coding world to see an evolution of services from narrowband (300-3400 Hz) speech to wideband (WB) or super-wideband (SWB) coders, which use a sampling rate of 16 kHz or 32 kHz for a frequency range of 50-7000Hz or 50-14000Hz respectively [1] [2].

The main WB/SWB speech codecs for mobile telephony, are AMR-WB for UMTS networks and VoLTE [2], EVRC-WB for CDMA networks [3], and EVS for VoLTE [4]. Several networks worldwide support AMR-WB on both UMTS and VoLTE, and EVS has also already been deployed on a handful of VoLTE networks.

However, there are big obstacles in deploying such codecs. Both terminal and network must be upgraded. On UMTS this may mean new radio-access bearers (RABs), as well as deployment of transcoder-free operation (TrFO) across the system (as otherwise conversion to PCM will cause the speech signal to revert to NB). Overall, operators will incur high costs to upgrade their network to a WB or SWB codec.

This has resulted in operators initially being very slow in deploying WB codecs. Currently adoption is finally getting faster for WB and SWB codecs, but often they are only offered on VoLTE only, as the cost of updating the 2G/3G infrastructure can be prohibitive.

3. WATERMARKING FOR WIDEBAND

3.1. Basic concept

Instead of using new WB/SWB codecs, a different approach has been proposed previously, using a WB codec which is fully backwards compatible with the existing NB infrastructure. [6]

The idea is to take a WB speech signal, and split it into a NB signal (typically 0-4 kHz), and a high-band (HB) signal (4-8 kHz). The HB signal can be coded efficiently using a small number of bits, as it typically contains much less information than the NB signal, and is highly correlated with the NB. The NB is coded with a standard NB codec, while the HB information is hidden in the NB bitstream using watermarking techniques. [7]

Here, we use bitstream watermarking. The bit patterns that can be transmitted by a codec, are constrained in a way that can be detected and decoded at the receiving end. Simple techniques exist, such as splitting a quantization table into two half-tables, and using indexes from the halftable corresponding to the bit of hidden data that is being sent. Provided the tables are well split, typically so that they both adequately cover the codeword space, information can be hidden with relatively little quality loss. [6]

3.2. Operating cases

To provide wideband speech, the decoder must receive the same bits that were sent by the encoder. This may not always be the case in a mobile telephony system, as the bitstream may be decoded to PCM and re-encoded in the network. As this is inefficient, and degrades quality, Transcoding Free Operation (TrFO) has been deployed on many UMTS networks, ensuring that the coded speech bitstream does not get re-encoded over the network. TrFO is becoming increasingly more common as it is more efficient in terms of capacity. Equivalent systems exist for 2G (TFO), and VoLTE calls can also avoid transcoding.

Overall, the watermarking codec provides the same user experience as the current NB codec when the bitstream is transcoded, or for mobile to landline calls. However when it is not, WB speech will be delivered.

3.3. Advantages/disadvantages over NB/WB codecs

Being backwards compatible with a pre-existing NB codec puts a lots of constraints on the design of watermarking codecs, and therefore they tend to have a slightly lower coding efficiency than a WB codec free of that constraint, at equal level of technology. However, in cases where only the watermarking codec can operate (because the conventional WB codec has not been deployed), this comparison is irrelevant, and the comparison should be with the existing NB codec.

A badly designed watermarking scheme may also introduce too much degradation in the legacy case. This is not acceptable, and the watermarking scheme must be such that its impact on NB quality is negligible.

The watermarking approach however has significant advantages: by not requiring any network changes, the only cost is that of deploying new software in the handsets, and ensuring the electro-acoustics work well with WB. It can also be used in complement to a WB/SWB codec, e.g. if a network only offers WB-SWB on VoLTE but not on the 2G and 3G parts of the network.

4. eAMR

4.1. Introduction

This paper introduces Qualcomm Technologies' new watermarking WB coder, called eAMR, for "enhanced AMR". eAMR is a watermarking coder based on the existing 3GPP AMR codec, a NB codec currently widely deployed in GSM and UMTS networks.

3GPP AMR is based on the ACELP paradigm, and offers a range of bit rates from 4.75 to 12.2, the top rate of 12.2 being identical to the older EFR codec. eAMR combines AMR with a HB coding scheme derived from the one used successfully in EVRC-WB, and a new dedicated watermarking scheme. Basic block diagrams of the encoder and decoder are shown in figures 1 and 2.



Figure 1: eAMR encoder



Figure 2: eAMR decoder

4.2. High band coding scheme

The high band coding scheme in eAMR is derived from that in EVRC-WB. The high-band is generated from the excitation of the NB codec, and only a few parameters are transmitted to the decoder to temporally and spectrally shape that signal. In EVRC-WB the high-band requires 0.8 or 1.35 kb/s depending on the frame type, but in AMR there is no frame type. Therefore a compromise is used, with 1.0 kb/s used all the time. An extra 4 bits are used for a CRC, leading to a total of 1.2kb/s for the watermark:

Parameter	Bits per 20ms frame
LSF	8
Frame gain	4
Gain shape	8
CRC	4
Total	24

Table 3:	Bit	allocation	for	the	high-band
----------	-----	------------	-----	-----	-----------

4.3. Watermarking scheme for the 12.2 mode

The watermarking scheme is based on constraining the positions of the pulses in the ACELP excitation. This is a classic watermarking technique for ACELP, and is detailed further in [6]. In EFR/AMR12.2, the fixed codebook is designed so that each 20ms frame (160 samples) is split into four 5ms frames of 40 samples structured as follows:

- Each subframe of 40 samples is split into 5 interleaved tracks, 8 positions per track
- 2 pulses and 1 sign bit per track, order of pulses determines second sign
- Pulse stacking is allowed
- Total: (2*3+1)*5=35 bits per subframe

Track	Pulses	Amplitudes	Positions
1	0, 5	±1,±1	0, 5, 10, 15, 20, 25, 30, 35
2	1,6	±1,±1	1, 6, 11, 16, 21, 26, 31, 36
3	2,7	±1,±1	2, 7, 12, 17, 22, 27, 32, 37
4	3, 8	±1,±1	3, 8, 14, 18, 23, 28, 33, 38
5	4,9	±1,±1	4, 9, 15, 19, 24, 29, 34, 39
777 1 1	4 10 1		

 Table 4: Pulse structure for AMR 12.2 fixed codebook

The watermark is applied by constraining the allowed pulse positions in a given track. For example, one bit per track can be hidden by forcing ($pos0^pos1$) & 001 to be the watermarking bit to be transmitted, where ANSI C operator notation is being used for the bitwise XOR (^) and bitwise AND (&). This constrains the pulse positions that are allowed, but hopefully the search algorithm will still find a good, although possibly sub-optimal, pulse combination. This would lead to 1 bit/track, e.g. 5 bits/subframe, 20bits/frame=1kbps of watermark.

Alternatively, forcing (pos0^pos1) & 011 to be 2 watermarking bits, would lead to a 2 kbps watermark.

The watermark is added by only carrying out the searches in the AMR fixed codebook (FCB) search which will decode into the right watermark. A problem however is that the main pitch pulse can be significantly affected, e.g. watermarking may prevent pulse stacking when it is needed. This has a bad impact on perceptual quality.

The solution used here consists in using an adaptive watermark, where the system predicts the likely position of the pitch pulse from the past transmitted parameters, does not touch the two tracks predicted to most likely contain the pitch pulse, and hides 2 bits in each of the three remaining tracks. The decoder is also able to recover this predicted position, so the positions do not need transmitting. This gives a total of 6 watermarking bits per 5ms subframe, i.e. a 1.2 kb/s watermark.

4.4. CRC scheme

A 4 bits CRC is used to detect errors in watermark. It has two main uses.

Firstly, it allows automatic detection of enhanced vs. legacy mode. This allows the eAMR decoder to know whether to decode the high-band bitstream or not. A 4-bit CRC is too short to give a reliable output, however a longer CRC would require a bigger watermark, which would degrade quality. In order to provide a reliable decision, the CRC is tracked over N frames (N=12) for a decision, and the bitstream is deemed to be enhanced if 7 or more correct CRCs are detected. WB output is produced in this case, otherwise it is NB only.

A secondary role is to detect errors. 4 bits is not enough to reliably determine all errors, but the NB bad frame indication (BFI) flag from the AMR channel coding should catch most frame errors. Some errors however may remain, e.g. due to bit errors being undetected (e.g. on Class C bits), The 4-bit CRC catches most such errors, and in practice works well on commercial 2G/3G/4G networks.

4.5. Other rates

AMR has 8 operating rates, ranging from 4.75 to 12.2 kbps. The intended operating point of eAMR is 12.2, as below that rate the advantage of WB coders over NB becomes smaller. However, the network may switch to lower rates in bad channel/network conditions. This is particularly true on 2G networks.

To avoid the risk of the bandwidth repeatedly switching between NB and WB, which would be very annoying for listeners, lower rates of eAMR are necessary to maintain WB speech at all times.

Therefore, eAMR has been implemented for all rates of AMR. Each rate has its own watermarking and quantization scheme, but still uses the same 4-bit CRC as the 12.2 scheme. The bit allocation per frame is as follows:

Rate	12.2	10.2	7.95	7.4	6.7	5.9	5.15	4.75
LSF	8	8	8	8	4	4	4	4
Frame Gain	4	4	4	4	4	4	4	4
Gain Shape	8	8	0	0	0	0	0	0
CRC	4	4	4	4	4	4	4	4
Total	24	24	16	16	12	12	12	12

Table 5: 1	High-band	bit :	allocation	for	all	rates	of	eAMR

5. PERFORMANCE

In order to characterize the performance of the eAMR coder, several subjective MOS tests were carried out at an external lab, according to the ITU P.800 ACR methodology. eAMR was compared to the AMR coder it is based on, and compatible with, as well as the AMR-WB standard, which is a dedicated WB coder, but is not compatible with AMR, and therefore requires an expensive new infrastructure for deployment.

A number of tests were carried out, only the most relevant results are presented here. Figure 6 shows the performance of eAMR 12.2 versus AMR-WB 12.65 and AMR 12.2. It is clear from the graph that eAMR provides nearly all of the benefit of going from AMR to AMR-WB, and is a very significant improvement over AMR. Indeed in this test, AMR-WB 12.65 and eAMR 12.2 are statistically equivalent with 95% confidence.



Figure 6: AMR vs eAMR vs AMR-WB, P800 ACR MOS

It can be noted that although AMR-WB 12.65 scores slightly better than eAMR 12.2, it is remarkable that the difference is so small, considering that eAMR has the huge disadvantage over AMR-WB of having to be compatible with standard AMR NB.

Figure 7 shows the legacy inter-op case, it is clear that the degradation introduced by the watermark is negligible and within the 95% confidence intervals. Therefore the deployment of eAMR should not degrade calls to legacy phones.



Figure 7: Legacy Inter-op case, eAMR 12.2 vs AMR 12.2, P800 ACR MOS

6. eAMR ON COMMERCIAL NETWORKS

eAMR was initially tested on call boxes using prototype handset devices, to test functionality and confirm quality. Further testing was carried out on various live commercial networks. It was found that eAMR performed satisfactorily on all the TrFO networks tested, and that the proportion of TrFO networks is high. Here is a non-exhaustive list:

- Europe: Vodafone in Germany, the UK and Italy, Telefonica in Spain, Telecom Italia Mobile in Italy, SFR and Orange in France, DT in Germany
- Asia: SKT in Korea, DCM in Japan
- USA: T-Mobile

Legacy operation was also tested on non-TrFO networks, with no issue. People listening to eAMR calls consistently expressed a clear preference for eAMR over AMR-NB, and generally felt that the performance of eAMR was comparable to that of AMR-WB.

7. CONCLUSION

A new way to deliver WB speech on existing NB networks was successfully developed and implemented, through watermarking a HB signal over an existing NB codec, namely AMR. Traditional methods of delivering WB/SWB speech require extensive changes to existing networks, however eAMR is a handset-only improvement. It does not require any changes to the network, and only needs a software upgrade in phones, provided the electro-acoustics of the phone are wideband capable. This enables an operator or OEM to deploy WB service rapidly and in a cost-efficient manner.

The proposed eAMR system has been shown to provide superior speech quality compared to conventional AMR, and to be a cheap and fast way to provide WB speech on current narrowband wireless networks. It can be deployed alongside existing WB/SWB networks, e.g. in the common case where the 2G/3G part of the network is NB while the WB/SWB offer is only on VoLTE.

eAMR is now commercially available on Qualcomm Technologies chipsets.

12. REFERENCES

[1] ETSI TS 126.090 (V9.0.0 2010-01), Adaptive Multi-Rate (AMR) speech codec; Transcoding functions (3GPP TS 26.090 version 9.0.0 Release 9).

[2] 3GPP TS 26.190, "Adaptive multi-rate wideband (AMR-WB) speech codec; Transcoding functions," 3rd Generation Partnership Project, Sept. 2012, version 11.0.0.

[3] 3GPP2 C.S0014-C vl.0, "Enhanced Variable Rate Codec, Speech Service Options 3, 68, and 70 for Wideband Spread Spectrum Digital Systems," <u>http://www.3gpp2.org/public-html/specs</u>.

[4] 3GPP TS 26.441, "Codec for Enhanced Voice Services (EVS); General overview," 3rd Generation Partnership Project, Dec. 2015, version 13.0.0

[5] http://www.voiceage.com/media/enablinghigh.pdf

[6] Geiser, B.; Vary, P.; , "High rate data hiding in ACELP speech codecs," Acoustics, Speech and Signal Processing, 2008. ICASSP 2008. IEEE International Conference on , vol., no., pp.4005-4008, March 31 2008-April 4 2008, doi: 10.1109/ICASSP.2008.4518532

[7] http://en.wikipedia.org/wiki/Steganography