FLOATING-POINT ADAPTIVE MULTI-RATE WIDEBAND SPEECH CODEC

Toni P. Nieminen
Nokia Research Center, Tampere, Finland

ABSTRACT

The Adaptive Multi-Rate Wideband (AMR-WB) speech codec algorithm has been selected for wideband speech coding in wireless and wireline services by both 3GPP and ITU-T. This paper describes an implementation of floating point Adaptive Multi-Rate Wideband (AMR-WB) codec that was approved by the Third Generation Partnership Project (3GPP) for multimedia applications.

1. INTRODUCTION

After the standardisation of Adaptive Multi-Rate (AMR) codec (3GPP TS26.073)[1] for 3GPP speech services, 3GPP wanted to widen the potential applications of AMR by providing low implementation complexity version of the codec. The specifications for conversational speech services in 3GPP are provided in fixed-point form using a library of so-called basic operations. Although this format is well suited to be used as a basis for a DSP implementation, it is not suitable for an implementation on a PC or other general-purpose processors. Modern general-purpose processors have very efficient floating-point units and thus 3GPP decided to implement more optimal version of the AMR codec using floating-point arithmetic. Encoder was changed to use floating-point arithmetic while decoder stayed in fixed-point arithmetic to simplify the interoperability testing. AMR floating-point codec (TS26.104)[2] was standardised in June 2000.

In March 2001 3GPP standardised AMR-WB speech codec for wideband speech services (TS26.173)[3]. In the beginning of year 2002, the AMR-WB speech codec was also selected by ITU-T for wideband speech coding in wireline services (G722.2).

Due to estimated wide deployment of 3G wireless systems, AMR-WB is also likely to be widely used in wireline packet based applications to eliminate transcoding between different standards. Multi-rate capability gives the user a possibility to adapt to the conditions of the surrounding medium. For example, the user can emphasise the quality of speech over video or give more bandwidth to the video stream by sacrificing some quality of speech. Normally wireline packet based applications operate on general-purpose processors and have multiple applications running simultaneously and thus fast execution of speech coding is required. The CPU utilisation of 3GPP reference fixed-point AMR-WB speech codec is so high that it makes it impossible to encode speech in real time using a low-end PC.

Work started in 3GPP to provide low complexity AMR-WB floating-point codec. The object of this new standard was to convert TS26.173 to use floating-point arithmetic and optimise execution in a way that real-time operation is possible in general purpose platforms.

2. AMR-WB CODEC DESCRIPTION

Most speech coding systems in use today are based on narrowband telephone-bandwidth speech from 200 to 3400 Hz. The AMR-WB speech codec utilizes a bandwidth ranging from 50 Hz to 7000 Hz improving intelligibility, naturalness and sense of presence. The AMR-WB speech codec utilizes the ACELP (Algebraic Code Excitation Linear Prediction) technology which is employed also in the AMR Narrowband and EFR speech codecs. The AMR-WB speech codec consists of nine speech codec modes with bit rates of 23.85, 23.05, 19.85, 18.25, 15.85, 14.25, 12.65, 8.85 and 6.60 kbit/s. AMR-WB includes also a background noise mode (Discontinuous transmission of speech, DTX). Bit-rate of this mode is 1.75 kbit/s.

AMR-WB speech codec provides superior voice quality over the existing narrowband standards. The clean speech quality of six highest modes is equal or better than ITU-T wideband codec G.722 at 64 kbit/s. The 12.65 kbit/s mode is at least equal to G.722 at 56 kbit/s. Two of the lowest modes 8.85 and 6.60 kbit/s are intended only for temporary use when severe error conditions occur. However, mode 8.85 kbit/s can still offer quality equal to G.722 at 48 kbit/s.

3. STANDARDISATION REQUIREMENTS

The floating-point AMR-WB codec specification (TS26.204)[4] is targeted to multimedia applications. The fixed-point specification 26.173 in 3GPP is still the only implementation allowed for the 3G-wideband-speech service. The target for the floating-point standardisation was to produce an ANSI C code that can be efficiently executed with general-purpose processors and, in particular, on the PC platform. The subjective quality of floating-point codec must correspond with that of the fixed-point codec. This requirement includes the quality when interoperating fixed- and floating-point codecs. In order to ensure the high and consistent quality of AMR-WB floating-point implementations, no modifications are allowed to be made to a C code and it is the only standard-conforming floating-point implementation allowed by the standard.

3GPP decided to use bit-exact decoder in the AMR-WB floating-point specification. It was seen that the target of fast
codec could be obtained also using fixed-point arithmetic in the decoder. Since the decoder consumes only a small fraction of the total processing effort of the AMR-WB codec, there is a negligible efficiency penalty in using a fixed-point decoder. There are many advantages of using a fixed-point decoder. The decoder performance can be verified using the AMR-WB test vectors. The fixed-point decoder also reduced the number of interoperability tests from three to one (floating-point encoder and fixed-point decoder). Using a fixed-point decoder also removed the need of testing the codec in error conditions.

To allow straightforward adoption of AMR-WB floating-point code, a new practical interface was added to the codec. This interface has all the functionality codec needs to operate efficiently, e.g. generation of transport stream and handling of different error conditions. The AMR-WB floating-point has optimised functions for using AMR-WB Interface Format 2 (3GPP TS 26.201)[5]. AMR-WB IF2 defines an octet-aligned frame format for the AMR-WB transport steam.

4. IMPLEMENTATION CONSIDERATIONS

The decision of using different arithmetic formats for encoder and decoder had to be taken into account when implementing the codec. ACELP coding is based on the linear predictive model of speech production, in which speech is generated by filtering an excitation signal. The excitation signal is constructed from the sum of two codewords, which are selected from the adaptive and fixed codebooks. The adaptive codebook contains delayed copies of the excitation signals. This enables a periodic excitation signal generated during voiced speech. The fixed codebook contains pulses, which contribute mainly during unvoiced speech, plosives and transitions. The contribution from each codebook in determined by its respective gain factor. The requirement that the AMR-WB floating-point encoder is interoperable with TS26.173 requires that the decoder have to be able to maintain the state of the encoder excitation signal memory for the next subframe bit-exactly. Even the slightest mismatch in decoder and encoder synthesis can cause problems and the states of the encoder and decoder can deviate and cause annoying effects in the decoder. The transmission errors can also cause similar deviations between encoder and decoder but they are corrected after error-free data is received and memories are thus resynchronised.

The mismatch is especially problematic with low energy voiced speech, where error in adaptive codebook cumulates rapidly. AMR-WB floating-point encoder solves this problem by providing bit-exact fixed-point excitation generation, even the analysis part is done in floating-point. This approach ensures encoder interoperability with fixed-point decoder.

5. VERIFICATION

The fact that a floating-point implementation is not bit-exact with TS26.173 raised a potential issue of interoperability and quality. Providing a standard floating-point implementation, which is verified to have good quality and interoperability with the fixed-point code, can solve this potential problem. In addition to giving a better quality guarantee this floating-point C code will also lower the threshold for implementation of the AMR-WB codec for non-3G platforms and thus improve interoperability of 3G multimedia and non-3G systems.

The extensive performance verification was performed prior to the acceptance of the specification. Codec was tested both subjectively and objectively. The 3GPP verification work included following items. [6]

- Verification of the format and correctness of the C-code
- Verification of subjective speech quality with respect to the existing AMR-WB fixed-point codec (subjective testing): clean speech, input levels, tandeming, background noise
- Verification of speech quality using objective measurements
- DTMF- and signaling tones
- Performance with music signals
- Special signals (in particular, non-speech signals)
- Check of overload performance
- Idle channel
- Complexity of the codec in different platforms
- Operation of the VAD and comfort noise
- Stability of the codec over time
- Bit-exactness of the decoder.

5.1. The bit-exactness testing of AMR-WB floating-point speech decoder

The AMR-WB floating-point speech decoder passes the AMR-WB test vectors. In addition, some further testing was performed to ensure the bit-exactness of the decoder and overall stability of the AMR-WB floating-point codec. This testing was done by first running large speech material through the AMR floating-point speech encoder. The output of the encoder was processed with the AMR-WB fixed-point speech decoder and with the floating-point speech decoder. Total amount of audio data was about 21 hours (>2 GB). One test contained very heavily clipped music with very high input level (-12 dBov). All the material was processed through all the 9 AMR-WB modes with DTX enabled.

5.2. Execution Time

Execution time of the floating-point codec was compared with the 3GPP fixed-point codec. Here the complexity of each version of the transcoder is measured based on how many ms it takes for the encoder to analyse all the input speech frames and simultaneously for the decoder to reconstruct all the output speech frames. Then the actual time spent processing all the frames is divided by the time of the input speech. This gives the average computational processing power (CPU load) per frame in real time. Time information was retrieved using a high-resolution performance counter. The high-resolution performance counter is hardware that provides high-resolution timing useful in improving the performance of applications. The Pentium has this kind of counter integrated. The following table has performance figures for the Pentium III 733MHz 256MB with Microsoft C compiler v6.0.
Listening Tests

Listening tests were performed as part of the 3GPP verification work. The verification test plan assumed that the performance of the AMR-WB floating-point code was compared with that of the official AMR-WB fixed-point code. All nine AMR-WB bit-rates were tested and a subset of the modes was tested with background noise recreation. The test was split into four experiments listed in table 1.

The Comparison Category Rating (CCR) method was used in testing. Listeners were presented with a pair of speech samples on each trial. The order of the AMR-WB fixed-point processed and floating-point codec samples is chosen at random for each trial. On half of the trials, the fixed-point sample is followed by the floating-point sample. On the remaining trials, the order is reversed. Listeners used the following scale to judge the quality of the second sample relative to that of the first:

The Quality of the Second Compared to the Quality of the First is:

-3 Much Worse
-2 Worse
-1 Slightly Worse
0 About the Same
1 Slightly Better
2 Better
3 Much Better

Table 1: Test plan for AMR-WB subjective quality verification.[8]

<table>
<thead>
<tr>
<th>Exp. No.</th>
<th>Title</th>
<th>Listening lab</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>CCR-test, Clean speech and input levels for the 5 modes (6.60, 8.85, 12.65, 18.25, 23.05 kbit/s)</td>
<td>Nokia</td>
</tr>
<tr>
<td>2</td>
<td>CCR-test, Clean speech and input levels for the 5 modes (6.60, 12.65, 15.85, 19.85, 23.85 kbit/s)</td>
<td>Ericsson</td>
</tr>
<tr>
<td>3</td>
<td>CCR-test, Background noise, Noise type: car noise</td>
<td>Nokia</td>
</tr>
<tr>
<td>4</td>
<td>CCR-test, Background noise, Noise type: babble noise</td>
<td>Ericsson</td>
</tr>
</tbody>
</table>

Comparison mean opinion scores (CMOS) for experiments are represented below. Experiment 1. included clean Finnish speech and three different input levels for modes 6.60, 12.65, 15.85, 19.85 and 23.85 kbit/s. Experiment 2 had clean Chinese speech and input levels for modes 6.60, 12.65, 15.85, 19.85 and 23.85 kbit/s. 192 votes per condition were cast by 24 listeners. Figure 3 shows CMOS values and confidence intervals for experiments one and two. Positive values signify preference for floating-point codec. Figure shows clearly that there is no difference in subjective quality between AMR-WB floating-point codec and AMR-WB fixed-point codec when clean speech is used.

Experiments 3 and 4 tested the performance in background noise. Experiment 3 used Finnish language and car noise, while experiment 4 used Chinese language and babble noise. Results in Figure 4 show that performance of fixed- and floating-point codecs is equal in most cases. Mode 15.85 kbit/s car noise has small preference for floating-point codec and mode 23.05 kbit/s babble noise has small preference for fixed-point codec.

Altogether, the results show that the performance of the AMR-WB floating-point is equal to that of the AMR-WB fixed-point. More information about the test procedure and results can be found from 3GPP technical documents: AMR-WB Floating-point verification test plan [7] and Subjective test results of the AMR-WB floating-point codec [8].

1 Listening tests 2 and 4 were performed by Ericsson/BIT research centre of digital communication technology (RCDCT) at Beijing Institute of Technology.
6. CONCLUSIONS

AMR Wideband floating-point codec is standardised by 3GPP and is the only floating-point realization allowed by the 3GPP standard. This ensures that the floating-point coded delivers always the same high quality wideband speech as corresponding fixed-point codecs. AMR-WB floating-point codec can be efficiently run on general processor platforms in real-time, and lowers the implementation threshold of AMR-WB codec.

7. REFERENCES


