High Definition Voice Rollout will Benefit all Mobile Users

Aram Falsafi
af@aramfalsafi.com

***Abstract*—As wireless operators begin to use voice quality as a differentiator, they will start to offer high definition voice services. With HD voice, the operator must pay close attention to voice quality degradations caused by tandem encoding and decoding – especially if a down-sampling to 64 kbps PCM is involved. In order to take full advantage of HD voice, operators must finally enable tandem-free operation (TFO) in their network equipment. Once TFO has been enabled, other equipment and services can take advantage of its capabilities, to improve voice quality, resulting in a general improvement in user experience for all subscribers. For example, true HD conference calls will be possible, and international mobile-to-mobile calls will suffer less from the adverse effects of multiple tandem vocoders.[1]**

Voice quality, transcoding, high definition voice, tandem-free operation, TFO (key words)

I. Introduction

Traditionally, voice quality was of secondary concern to mobile subscribers. Initially, users were happy simply to place and receive voice calls from their automobiles. Later, with improvements in RF electronics and battery technology, "car phones" became "mobile phones" small enough to be carried on the person. With the advent of 2.5G and 3G technologies, the mobile phone became a mobile, always-on, always-connected voice/data device.

Throughout this time, over-the-air bandwidth continued to be a scarce resource. Therefore, second generation, and even third generation mobile standards incorporated aggressive voice compression standards, and as a result, users became accustomed to low voice quality.

Up until now, mobile operators have distinguished themselves through features such as geographic coverage, handsets with rich features, and low subscription cost. In the U.S., Verizon's recent ads touting its coverage, AT&T's exclusive contract with Apple, and Sprint's announcement that it will continue to offer unlimited data plans are some examples of the above.

As mature markets reach saturation, and as operators find ways to offer devices and coverage comparable to their competitors, some have started to look at voice quality as a way to distinguish themselves and gain a competitive advantage, even if for a short period of time. Initial experiments have shown great satisfaction on the part of subscribers [17].

At the same time, Skype and other IP telephony applications have given the general public a taste of the benefits of high-definition audio. Thanks to Skype's use of HD vocoders, calls placed from one Skype user to another exhibit much higher voice quality than landline-to-landline calls, which have been the so-called "gold standard" for voice communication until very recently.

[1] The author is a Network Consultant with Tellabs professional services group

II. AMR-WB, the HD Voice Standard for Mobile

High definition voice, standardized in 2001 by 3GPP (3rd Generation Partnership Project) and known as Adaptive Multi-Rate Wideband or AMR-WB [12, 13], is an opportunity for wireless operators to leapfrog over landline telephony, offering subscribers voice quality comparable to Skype-to-Skype calls.

In traditional telephony, the input voice is bandpass filtered to remove components below 200 Hz, and above 3400 Hz. The filtered signal is then sampled at 8,000 samples per second, and quantized at 8 bits per sample, to generate the well-known 64 kbps Pulse Code Modulation (PCM) signal.

In AMR-WB, the bandpass filter is set to 50-7000 Hz, and the filtered signal is sampled at 16,000 samples per second. Subjective studies have concluded that addition of the 50-200 Hz band at the low end contributes to increased "naturalness" and gives the impression of being in the same room with the speaker, while extension of the high end from 3400 Hz to 7000 Hz provides higher intelligibility [18].

The AMR-WB codec specifies nine bit rates, from 6.6 kbps up to 23.85 kbps, and is based on an algebraic code-excited linear prediction (ACELP) algorithm, similar to the most common narrowband mobile vocoders. However, thanks to its wider frequency range, AMR-WB provides much better voice quality than *the best narrowband vocoders operating at comparable or even higher data rates*. For example, AMR-WB at 8.85 kbps out-performs traditional narrowband AMR at 12.2 kbps in subjective voice quality tests [20].

The same vocoder has also been standardized for other mobile services, such as

- packet-switched conversational multimedia [5]
- multimedia messaging service [6]
- packet-switched streaming service [7]
- IMS messaging and presence [8]
- multimedia broadcast multicast service [9]

III. The Interoperability Challenge

AMR-WB is easy to implement when a call is placed from one mobile to another mobile, both of which are part of the same operator's network. However, given the number of companies providing voice services – fixed line, mobile, over-the-net, corporate PBX/Centrex, voicemail boxes, conference bridges, etc. – it will be years before all interconnections are based on standards that support high definition voice.

In the meantime, many calls that originate as HD will pass through interconnections that are based on traditional 64 kbps PCM transport. The near-end switch has no easy way to determine if the far end device supports AMR-WB. It will therefore have to filter the voice signal to 200-3400 Hz and sample it at 8000 samples/second. The down-sampling will negate all advantages of high definition voice described above, even if the signal could later be converted back to AMR-WB.

Since the AMR-WB signal, even at its highest data rates, would easily fit into a 64 kbps PCM channel, this down-sampling would be unnecessary, if there were a way for the near-end gateway to query the far-end gateway as to its wideband capabilities.

Fortunately, there is a solution to this problem, based on an extension to a standard that was developed years ago, in order to address a different problem – the rapid degradation in voice quality that results from passing a voice stream through various sets of narrowband encoders and decoders. First let us review the original standard, before describing the extension.

IV. TANDEM-FREE OPERATION

The negative effect of tandem vocoding on voice quality has been known since the 1990's, long before standardization of high definition voice. As a result, both the GSM and the CDMA (IS-95) standards developed methods for in-band signaling to reduce the number of tandem vocoders when mobile to mobile calls pass through a PCM channel [1, 2]. This is known as Tandem-Free Operation (TFO). At call setup time, a 64 kbps PCM channel is established between the switches terminating the compressed narrowband voice connection from each mobile. The two switches then use in-band signaling to exchange information about the mobiles' vocoder capabilities, select a common vocoder, and inform the two mobile devices. The narrowband voice can then bypass the network-side encoders and decoders, carried over the lowest significant bit(s) of the 64 kbps PCM channel.

The improvement in voice quality as a result of TFO can be easily quantified. Figure 1, reproduced from [16], shows the effect of tandem vocoding on voice quality for different narrowband mobile vocoders. Voice quality was measured using the standard Perceptual Evaluation of Speech Quality (PESQ) algorithm [4]. The vocoders studied were standard 3GPP vocoders: Full Rate, Half Rate, Enhanced Full Rate, and Adaptive Multi-Rate at 4.75, 10.2, and 12.2 kilobits per second.

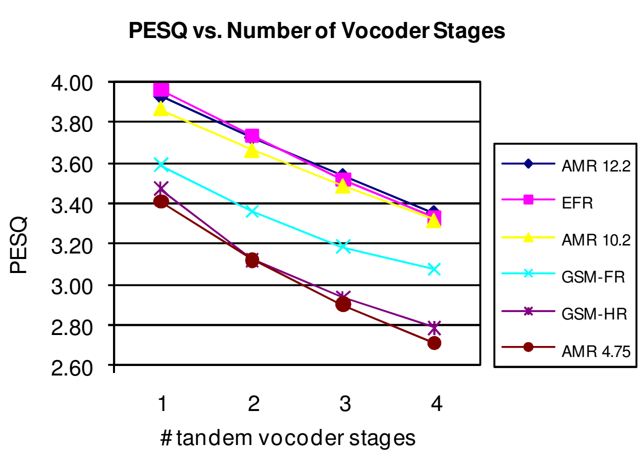

Figure 1 – Effect of tandem vocoders on voice quality for narrowband voice

Figure 2 shows a call with 3 tandem encoder/decoder pairs. The second tandem vocoder pair are two Digital Circuit Multiplication Equipment (DCME) devices, found on many international circuits where voice compression is needed due to the high cost of bandwidth. For simplicity, voice traffic in only one direction is shown. The number of tandem vocoders can further increase if, for example, the wireless operator uses media gateways between the Radio Network Controller (RNC) and the core network, and configures these gateways to use a vocoder other than the standard G.711 at 64 kbps. This is a common practice, especially in emerging markets.

It is worth mentioning that Figure 1 only shows the effect of tandeming vocoders of the same type, also known as self-tandeming. The degradation in voice quality is even more severe if different vocoders are placed in tandem [11, 15].

V. APPLICATION OF TFO TO HIGH DEFINITION VOICE

As mobile voice traffic is migrated over to packet switched networks, HD voice can be carried end to end without the need for transcoding. Examples include calls in a full-IP network and calls placed between two subscribers connected to the same RNC, or connected to RNCs that have an IP connection between them. This is standardized by 3GPP as Transcoder-Free Operation (TrFO) and uses out-of-band signaling [17].

However, these scenarios are by no means universal. There are many cases where the voice call will pass through a 64 kbps, circuit-switched link. Examples include:

- International calls
- Calls between different mobile operators in many emerging markets
- Calls to IP-PBX systems that interface to the public switch telephone network with T1/E1 connections
- Calls to many conference bridge numbers

With TFO enabled, the two mobile networks will negotiate a codec rate that both mobiles can support. They will then bypass their network-side encoder and decoder units. AMR-WB calls can then cross PCM links, carried using a subset of the bits in each PCM sample.

Therefore, in the short term, mobile operators who want to offer HD voice to more than a limited subscriber base have no choice but to enable TFO in their networks.

In the case of 4th generation networks (specifically Long Term Evolution or LTE), the plan is for full-IP connectivity with voice carried over IP Multimedia Systems (IMS).

However, LTE/IMS will not be universally available any time soon. According to one LTE equipment manufacturer, quoted in a 2010 market study on voice over LTE, many mobile operators "say they see use of GSM networks for voice for quite a while, even 10 to 20 years" [21].

VI. EFFECT OF TANDEM VOCODING ON DELAY

Most mobile vocoders use a block size of 20 milliseconds. Many, including narrowband AMR, also use a look-ahead buffer of 5 ms.

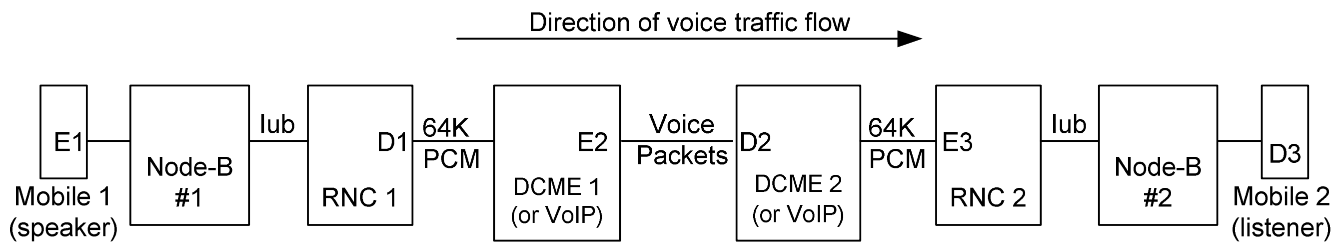

Figure 2 – Three-stage tandem vocoders with three encoders (E1, E2, E3) and three decoders (D1, D2, D3)

This means that every encoder must collect 20-25 ms of data before applying the vocoder algorithm, which itself introduces a few ms of delay. At the decoder, the entire 20 ms frame must arrive before the decoding process can begin. This introduces another delay, which could be significant or not, depending on the speed of the interface and processor capabilities. Other sources of delay include interleaving, time division multiplexing, and queues and buffers in the transmission path.

Figure 3, reproduced from ITU recommendation G.114 on speech delay [19], shows that some users start to express dissatisfaction with speech quality when delay exceeds 150 ms. In general, a one-way delay of 250 ms is considered the absolute maximum that should be budgeted [11].

While delay is a function of many factors, and varies from network to network, having more than two pairs of encoders and decoders can easily bring the delay close to the ideal limit of 150 ms. Every tandem vocoding stage that can be avoided reduces the overall delay by at least 25-50 ms.

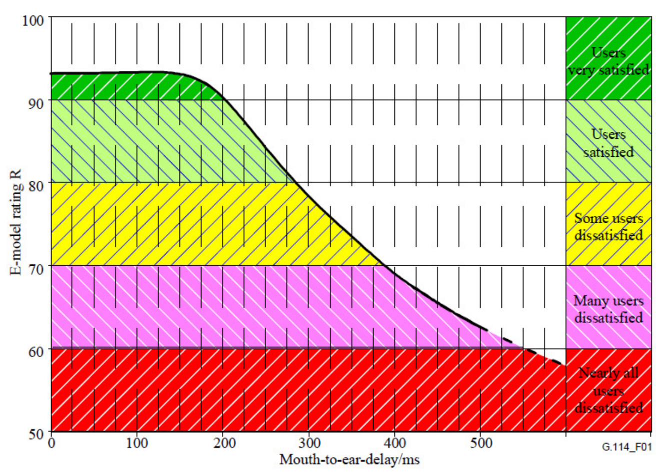

Figure 3 – User satisfaction as a function of one-way speech delay

VII. The Economic Case for Tandem-Free Operation

Another reason to avoid tandem vocoding is the cost of dedicated processing needed for real-time encoding and decoding of speech. In Figure 4, the only processing elements that are absolutely required are E1 and D3, the encoder and decoder that are already built into the two mobile handsets. The price of these is already included in the cost of handsets, which are either paid for fully by the subscriber, or subsidized by the service provider, in exchange for a long-term contract.

Since real-time voice compression/decompression is very processor-intensive, the equipment manufacturer can realize significant cost savings by minimizing the need for this functionality in network equipment – costs savings that can be passed to the operator.

An AMR-WB vocoder requires twice the processing of a narrowband AMR vocoder [17]. Therefore, introducing AMR-WB without widespread use of TFO would require roughly a doubling of processing resources in the operator's Transcoder Rate Adapter Unit (TRAU). On the other hand, TFO, even when applied to narrowband voice calls, will free up vocoder resources that can then be allocated to AMR-WB calls.

VIII. Added Benefits of TFO in the Network

The good news is that enabling TFO in the network will bring with it additional benefits for the operator's entire subscriber base. Two of the benefits, namely improved voice quality and lower delay, were described above.

Once TFO is enabled, other equipment interfacing to the mobile network over circuit-switched connections can also take advantage of this feature by essentially pretending to be another TFO-capable mobile switch. This will allow the third-party equipment to receive the voice signal as it was originally encoded by the mobile handset, without the first-stage decoding (D1 stage in the above example).

Since TFO signaling is in-band on the 64 kbps PCM channel, any device connected to this channel can listen for TFO handshaking messages and respond accordingly. The procedure can be quite simple, and is described in [3].

In some applications, such as DCME on long-distance international links, the primary beneficiary will be narrowband voice users. Anyone who places frequent international calls to mobile phones, especially in emerging markets, knows that voice quality can vary from barely acceptable to completely unintelligible. This is in part due to the third tandem vocoder on the international link, which can use quite aggressive compression rates, especially during peak hours (E2/D2 in Figure 2 above).

In other applications, such as conference bridges, HD voice users will benefit the most. With TFO, the conference bridge can signal the mobile network to not convert the voice stream into PCM. If most of the conference participants are using HD voice capable handsets, these participants will enjoy a much higher speech quality while one of them is speaking.

Figure 4 shows a conference bridge where TFO is used for mobile devices with AMR-WB codecs. For simplicity, only one speaker and one listener are shown. At call setup time, the conference bridge uses TFO negotiation to force all AMR-WB capable mobiles to the same codec rate. During the conference call, the speaker's voice is encoded with a AMR-WB codec at his or her mobile (encoder E1). This AMR-WB signal is then carried all the way to the mobile device of every listener, where it is decoded (decoder D1). There is no tandem vocoding, and more importantly, the speech is never down-sampled to the 8000 samples/second PCM rate. Of course, the conference bridge will have to decode the speech down-sample it to 8000 samples/second, convert it to 64 kbps PCM, and distribute this copy to all callers who do not have a AMR-WB mobile.

A third application, and one where TFO can benefit both narrowband and HD voice users, is speech recognition. TFO will improve the reliability of speech recognition when the speaker is on a mobile handset [14].

These three cases are just some examples of the benefits of enabling TFO in the mobile network and adding this capability to third party devices.

Unfortunately, this has not been possible up until now, because most operators have not yet enabled TFO in their networks – the voice quality on most local and national calls had been "good enough" and the marginal benefits did not justify the extra cost and complexity.

IX. Conclusion

With the introduction of high definition voice in the form of AMR-WB, mobile operators will have no choice but to enable tandem-free operation in their networks. Once TFO has been enabled, even subscribers who don't have AMR-WB handsets will benefit.

Narrowband mobile-to-mobile calls will experience improved voice quality and lower delay.

Manufacturers of third party equipment such as DCME devices and media gateways, conference bridges, speech recognition platforms, and others can incorporate TFO into their products and offer an improved user experience.

Finally, the mobile operators will benefit economically by making more efficient use of their existing codec hardware.

X. References

[1] 3GPP TS 28.062, "Inband Tandem Free Operation (TFO) of Speech Codecs; Service Description; Stage 3," V9.0.0, December 2009, http://www.3gpp.org/FTP/Specs/html-info/28062.htm.

[2] CDMA Tandem Free Operation, 3GPP2 A.S0004-B v2.0, August 5, 2002, http://www.3gpp2.org/Public_html/specs/A.S0004-B_v2.0.pdf.

[3] Falsafi, A., "Method for Reducing the Number of Vocoders in a Wireless Network," U.S. Patent No. 7,023,819 B2, 4 April 2006.

[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," International Telecomm. Union, Geneva, Switzerland, February 2001.

[5] 3GPP TS 26.235, "Packet Switched Conversational Multimedia Applications; Default codecs," V9.0.0, December 2009.

[6] 3GPP TS 26.140, Multimedia Messaging Service (MMS); Media Formats and Codecs," V9.0.0, December 2009.

[7] 3GPP TS 26.234, "Transparent end-to-end Packet-switched Streaming Service (PSS); Protocols and codecs," V9.4.0, September 2010.

[8] 3GPP TS 26.141, "IP Multimedia System (IMS) Messaging and Presence; Media Formats and Codecs," V9.0.0, December 2009.

[9] 3GPP TS 26.346, "Multimedia Broadcast/Multicast Service (MBMS); Protocols and Codecs," V9.4.0, September 2010.

[10] Redding, C., DeMinco, N., Lindner, J., "Voice Quality Assessment of Vocoders in Tandem Configuration," NTIA Report 01-386, April 2001, available from http://ntia.its.bldrdoc.gov/pub/ntia-rpt/01-386/01-386.pdf.

[11] Gibson, J., "Speech Coding Methods, Standards, and Applications," IEEE Circuits and Systems Magazine, Vol. 5, No. 4, 4Q2005, pp. 30-49.

[12] 3GPP TS 26.190, "Adaptive Multi-Rate - Wideband (AMR-WB) speech codec; Transcoding functions" V9.0.0, December 2009, available from http://www.3gpp.org/FTP/Specs/html-info/26190.htm.

[13] 3GPP TS 26.194, "Adaptive Multi-Rate Wideband Speech Codec; Voice Activity Detector (VAD)," V9.0.0, December 2009, available from http://www.3gpp.org/FTP/Specs/html-info/26194.htm.

[14] Gómez, A., Peinado, A., Sánchez, V., Rubio, A., "Recognition of Coded Speech Transmitted over Wireless Channels," IEEE Transactions on Wireless Communications, Vol. 5, No. 9, Sept. 2006, pp. 2555-2562.

[15] Ramo, A., Toukomaa, H., "On Comparing Speech Quality of Various Narrow- and Wideband Speech Codecs," Eighth International Symposium on Signal Processing and Its Applications, Sydney, Australia, August 2005, pp. 603–606.

[16] Falsafi, A., "Improving Voice Quality in International Mobile-to-Mobile Calls," IEEE International Symposium on Personal, Indoor, and Mobile Radio Communications, Cannes, France, September 2008.

[17] Birkehammar, C., Bruhn, S., Eneroth, P., Hellwig, K., Johansson, S., "New High-Quality Voice Service for Mobile Networks," Ericsson Review, No. 3, 2006, available for download from www.ericsson.com/ericsson/corpinfo/publications/review/2006_03.

[18] Bessette, B., Salami, R., Lefebvre, R., Jelinek, M., Rotola-Pukkila, J., Vainio, J., Mikkola, H., Jarvinen, K., "The Adaptive Multi-Rate Wideband Speech Codec (AMR-WB)," IEEE Trans. Speech and Audio Processing, vol. 10, no. 8, Nov. 2002, pp. 620–636.

[19] ITU-T Recommendation G.114, "One-way transmission time," International Telecomm. Union, Geneva, Switzerland, May 2003.

[20] Ojala, P., Lakaniemi, H., Lepanaho, H., Jokimies, M., "The Adaptive Multirate Wideband Speech Codec: System Characteristics, Quality Advances, and Deployment Strategies," IEEE Communications Magazine, Vol. 44, No. 5, May 2006, pp 59-65.

[21] "Voice Over LTE: More Pitfalls than Promise for Now," Heavy Reading 4G/LTE Insider, Vol. 1, No. 4, October 2010, available from http://www.heavyreading.com.

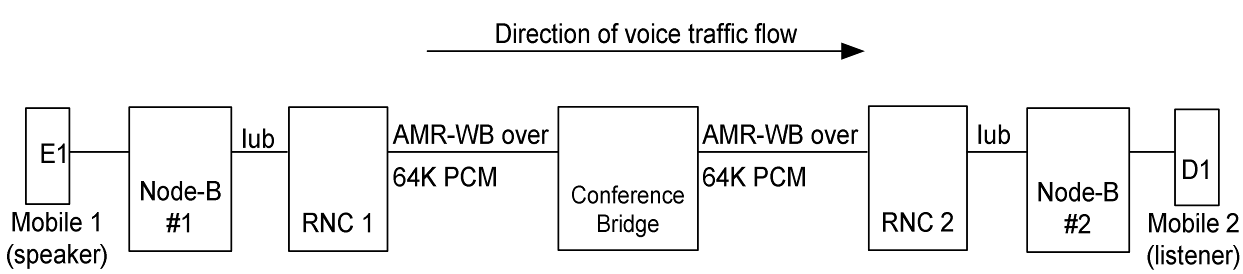

Figure 4 – Conference call with two AMR-WB callers and TFO enabled