Trends & Technical Challenges in Conversational Voice Services

A. Ryan Heidari
Qualcomm Incorporated, 5775 Morehouse Drive, San Diego, CA 92121
rheidari@qualcomm.com

ABSTRACT

Technologies continuously evolve to meet the challenging needs of wireless communication. The ever growing demand for wireless communication requires greater capacity and higher network efficiencies. Spectral efficiency is a key driver of the economics of voice and data services as we approach the rollout of 3G network. In this paper we discuss the technology trends and challenges associated with conversational voice services in the context of fixed mobile conversion. We discuss evolution in cellular modem technology, switching technology, and speech coding technology. The combination of voice over internet protocol and wireless communication promises to revolutionize the whole telecommunication market.

Index Terms— Speech Coding, Telephony, Wireless voice services, Multimedia communication, Land mobile radio data communication, Internetworking, Voice over IP.

1. INTRODUCTION

Wireless operators worldwide are witnessing an exponential increase in their subscribers' growth and demand for conversational voice services remains very strong. In 1990 there were just over 10 million mobile subscribers around the world, by the middle of the nineties this figure had increased by 10-fold, and by the end of 2004, there were over 2 billion mobile subscribers accounting for roughly half the world's telephone lines. This growth is unparalleled in the history of technology adoption, even compared with radio, television or the internet proliferation.

As user demand for data-intensive applications continues to grow, speech codecs with better coding efficiency provide network operators the ability to accommodate their existing network for voice services. Furthermore, we see a proliferation of Internet Protocol (IP) phones as we approach the rollout of 3G wireless handsets. This technology evolution along with fixed mobile conversion or convergence (FMC) (which is intended to provide telecommunication applications and services independent of the access network, mobile or fixed) has created a tremendous opportunity to provide better than toll quality voice services along with rich media call features in the next-generation fixed mobile telephone network.

Conversational voice services still remain the prime source of revenue for the wireless operators while other data services are slowly catching up. Fig. 1 shows the percent of market share as a function of revenue for non-conversational voice services such as premium and standard data services. For example, in 2006 83% of total revenue still belongs to conversational voice services while the remaining 17% is accounted for other wireless data services. Among the data services messaging such as short messaging service (SMS), multimedia messaging service (MMS) and instant messenger (IM) are still the key driver for non-voice services as compared with the premium services such as gaming, audio and video application. For example, in 2006 the revenue from premium data services was about 9% while the messaging revenue was about 35%. The remaining revenue from data services comes from minutes/mbytes in high speech data services.

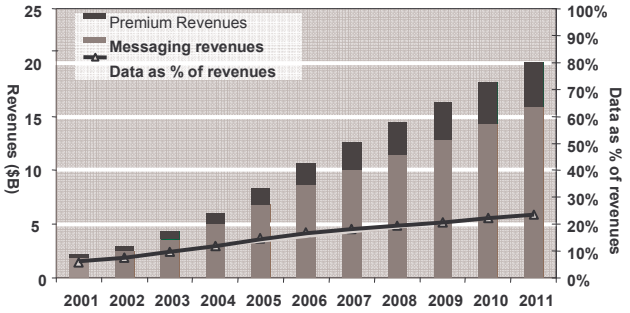

Fig. 1 Percent of data services and their associated revenues

Operators continue to expand their voice services by enhancing their voice services by offering high fidelity wideband voice quality within their network, transitioning to media rich IP calls and attracting more landline subscribers by offering cheaper long-distance and in-network calls.

2. EVOLUTION OF WIRELESS MODEM TECHNOLOGY

The cellular modem design has gone through major technology advancement during the past decade as shown in Fig. 2. Such technology evolution was initially driven by demand for higher voice channel capacity while coping with the subscribers' growth and working around inadequate spectral bandwidth allocation. Then more recently the concentration has been shifted toward enhancing the high speed data rate capability for more advanced multimedia services. The first generation (1G) cellular system was introduced with advanced mobile phone system (AMPS) using analog voice as prime application with very limited

data application. Soon after that the technology evolved into two different digital camps, global system for mobile communication (GSM) and code division multiple access (CDMA), by offering the second generation (2G) cellular system with more spectral coding efficiency and better digital voice quality than AMPS. It didn't take much time for the technology to offer higher data rate capabilities (while maintaining backward compatibility) with GSM release 97 (Rel-97) general packet radio service (GPRS), Rel-99 enhanced data for GSM evolution (EDGE), and CDMA evolution data optimized (1xEV-DO). Such 2.5G cellular system was designed to better handle moderate speed data services such as SMS and internet communication service such as email.

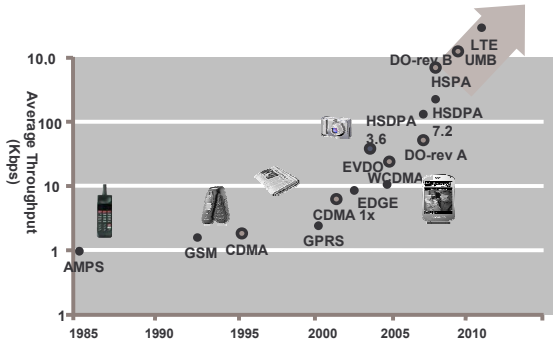

Fig. 2 Trends in wireless cellular modem technology

With the emergence of universal mobile telecommunications system (UMTS) and wideband code division multiple access (WCDMA) technology, the third generation (3G) was introduced with higher data rate transfer and more voice capacity. High speed downlink packet data (HSDPA) grew from Rel-4 1.8 Mbps to Rel-6 with up to 14.4 Mbps. The CDMA equivalent, 1xEV-DO Rev A with better QoS parameters for real-time services, offers maximum down-link speeds of 3.1 Mbps. High-Speed Packet Access (HSPA) evolution or better known as Rel-7 reaches speed of 42 Mbps in downlink and 11 Mbps in uplink.

Fourth generation (4G) cellular system introduces long term evolution (LTE) or Rel-8 and CDMA ultra mobile broadband (UMB), both based on orthogonal frequency division multiple access (OFDMA); they can provide high speed data in excess of 100 Mbps in downlink and 50Mbps in uplink. Along with modem technology evolution and their respective coding efficiencies to achieve higher data rates, the total number of voice calls per given 10 MHz cell has been improved as well [4]. Table 1 shows total number of simultaneous calls supported for different speech codecs in GSM, WCDMA and CDMA network. GSM EFR and AMR 12.2 kbps codecs (see Section 4) are shown with frequency reuse factors of 3/9 and 1/1 respectively. (Frequency reuse factor refers to the rate at which a frequency can be reused in a network.) All other technologies are using a frequency reuse factor of 1/1. As you can see the number of calls has been increased by more than two folds in WCDMA and CDMA-1x network as compared with the legacy GSM and CDMA network. There is even more capacity gain possible with codecs operating at lower peak or average data rates while allowing to tradeoff for lower voice quality. Furthermore, the number of packet switched voice over internet protocol (VoIP) calls in HSPA and 1xEV-DO Rev A (DOrA) have been further increased as compared with their respective circuit switched network.

GSM		WCDMA		HSPA-VoIP	
EFR 3/9	41	AMR12.2	120	AMR-Rel6	196
AMR 3/9	74	AMR5.9	160	AMR-Rel7	380
AMR 1/1	103				
CDMA		**CDMA-1x**		**DOrA -VoIP**	
EVRC	147	EVRC	245	EVRC	308
		EVRC-B	343	EVRC-B	336

Table 1 Number of calls per sector for different technologies

3. EVOLUTION FROM CIRCUIT SWITCHING TO PACKET SWITCHING

Telecommunication network has gone through a major architectural overhaul since its original inception many decades ago. The network was originated as analog wireline public switched telephone network (PSTN) by passing the analog speech signal over a long haul of copper wire in a peer-to-peer configuration. The voice signal had to be bandlimited to about 200-3400 Hz to accommodate the bandwidth characteristic of copper wire. This narrowband voice quality is sometimes referred to as "toll quality". Later with objectives to better utilize and manage the circuit switched call flow in core network, a 64 kbps digital voice channel was introduced. This digital voice channel was designed to carry phone calls from calling party to called party using basic digital circuit switched PSTN.

Subsequently, the wireless telecommunication core network pursued similar dedicated channel circuit switched concept as PSTN network. But they had to use more compressed digital voice channel structure with higher speech coding efficiency to overcome the band-limited wireless transmission. Later on with the introduction of IP core network the concept of packet switched transport was introduced. In the circuit switched case the compressed vocoder frames of data are directly modulated and transmitted over the air by maintaining the dedicated channel characteristics without any need for additional transport layers. In a sense the QoS factors such as delay and transmission frame losses are well controlled. In contrast, the packet switched case relies on the best effort transmission using IP core network along with internet engineering task force (IETF) transport layers. Although packet switched network is more robust for expansion and costs less than circuit switched network but it is not as

bandwidth efficient due to its additional overhead associated with transport layers. The QoS factors such as delay and transmission packet losses are not generally well controlled partly due to the best effort transmission.

In packet switched model the vocoder packet is converted to data packet when transported over IP network using real time protocol (RTP), user datagram protocol (UDP) and finally internet protocol (IP) hearders [6]. Without using header compression technologies such as robust header compression (ROHC) such transport layers may introduce up to 80 bytes of overhead per packet.

4. SPEECH CODEC TECHNOLOGY EVOLUTION

The digital speech processing technology has gone through an elaborate architectural design challenges during the past few decades. Designers have introduced many innovative techniques to further improve the overall coding efficiency of speech codecs and lowering their data rates while preserving the desirable voice quality. Furthermore, the real-time implementation of these techniques has been possible with enhancement in digital signal processors (DSP) and increase in their overall computational power. Such progression in codec design has created many different tiers of speech codecs, each with their own unique characteristics that are optimized for specific application. Fig. 3 shows an ensemble of such codecs along with their peak or average data rates and their respective standardization body. In addition to the data rates, there are many other design factors that contribute to the technical trends and challenges among these codecs. For wireless application in particular, which is the focus of this paper, other factors such as voice quality in clean and noisy condition due to its mobility usage, robustness against higher packet losses caused by combined wireless and wireline transmission channels, methods to complement against increase in throughput delay, and finally lower complexity are important factors.

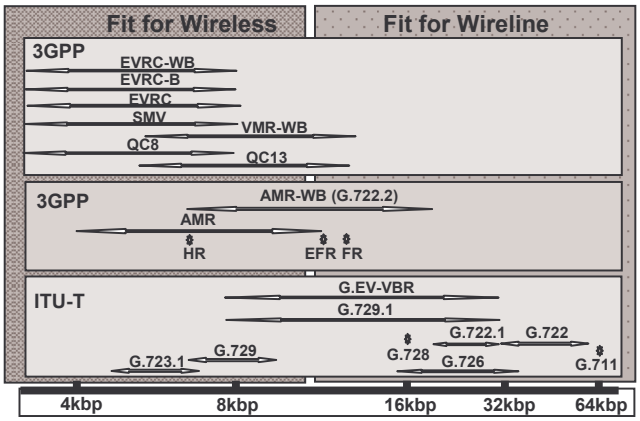

Fig. 3 Landscape of wireline and wireless speech Codecs

In this section we consider a few standardization bodies and their respective vocoder roadmaps along with their key design objectives. There are two families of codecs for conversational voice services; the narrowband (NB) and wideband (WB). The NB family of codecs are sampled at 8 kHz with bandwidth around 300-3200 Hz. The WB codecs are sampled at 16 kHz with bandwidth around 50-7000 Hz. The WB codecs require higher data rates but provide better voice quality with improved intelligibility and naturalness.

International Telecommunication Union's Telecommunication sector (ITU-T) has standardized a series of speech codecs [7], mainly for wireline broadband and fixed network. The technology evolution in ITU-T has been to progressively lower bit rates while maintaining speech quality, both in narrowband and wideband cases. The ITU-T NB family of codecs consists of G.711 (1988), logarithmic pulse-code modulation (PCM) at 64 kbps to match the frame structure of digital PSTN and integrated services digital network (ISDN); G.726 (1990), adaptive differential pulse code modulation (ADPCM) at 16-40 kbps; G.728 (1995), low delay code excited linear prediction codec (LD-CELP) at 16 kbps; G.729 (1996), conjugate-structure algebraic-code-excited linear prediction (CS-ACELP) at 8 kbps with extension to 6.4 and 11.8 kbps; and finally G.723.1 (1996), operating at 6.4 and 5.3 kbps. The codecs with lower and variable peak data rates were designed to fulfill the requirement of asynchronous transfer mode (ATM) and IP network. The WB family of codecs consists of G.722 (1986), subband based codec where the input signal is split into two bands and separately encoded using ADPCM at total bit rates of 48-64 kbps; G.722.1 (1999) at 24-32 kbps; and finally G.722.2 (2002) which is also known as adaptive multi-rate wideband (AMR-WB) [2] at 6.6-23.85 kbps. These WB codecs can broaden the conventional voice services in other applications such as teleconferencing and video telephony. Furthermore, ITU-T has introduced two new codecs; G.729.1 (2006) [8] and G.EV-VBR (2008) where the core of the codec is extended up to 32kbps in an embedded multi-layer fashion to further enhance the quality of audio and music signals and to improve speech quality under packet loss. These embedded coders are scalable in bit rate and speech bandwidth, supporting both narrowband and wideband speech.

The third generation partnership project (3GPP) [1] has standardized a series of codecs for GSM and UMTS networks. The challenges here are on ways to optimize the source and channel coding based on GSM traffic channel of either 11.4 kbps or 22.8 kbps. Furthermore, the source coding bits are classified into class A and class B. Class A bits are considered more important than class B and well protected against transmission errors to deliver a more consistent voice quality. This family of codecs consists of GSM full rate (FR) (1987), regular pulse excitationlong term prediction (RPE-LTP) at 13 kbps;, enhanced full-rate (EFR) (1995) at 12.2 kbps; GSM half-rate codec (1994), vector-sum excited linear prediction (VSELP) at 6.1 kbps; adaptive multi-rate (AMR) (1999), algebraic code excited linear

prediction (ACELP) at 4.75-12.2 kbps; and finally adaptive multi-rate wideband (AMR-WB) (2001) [2] at 6.6-23.85 kbps. AMR or AMR-WB codec uses multiple fixed rates, whereby the desired rate is selected in an adaptive fashion to tradeoff between voice quality and cell coverage. All 3GPP standard codecs are fixed-rate codecs and rely on use of voice activity detection (VAD), discontinuous transmission (DTX), and comfort noise generation (CNG) for detecting, coding and regeneration of silent and background noise periods respectively in full-duplex conversational services.

The third generation partnership project 2 (3GPP2) [3] has standardized a series of codecs for CDMA network. Contrary to 3GPP all 3GPP2 standard codecs are source controlled variable rate codecs, to take full advantage of CDMA physical layers and its associated rate configuration. CDMA spread spectrum technology voice capacity is calculated based on average transmitted power rather than peak power and as such can benefit from a variable rate codec to further reduce the average data rates and consequently lower the transmission power to achieve higher voice capacity. There are two distinct multiplex sub-layers defined in the standard; rate set 1 at 8.55 kbps and rate set 2 at 13.3 kbps. The codec peak data rate is based on one of these two rate sets but the average data rate can be lowered based on percent usage of one of the four different rate configurations; full-rate, half-rate, quarter-rate and eighth-rate. The proper rate configuration is selected based on source controlled waveform classifier and voice activity. 3GPP2 NB standard codecs consist of Qualcomm 8k (QC8) (1994) at a peak data rate of 8.55 kbps, Qualcomm 13k (QC13) (1996) at a peak rate of 13.3 kbps, enhanced variable rate codec (EVRC) (1998) at a peak rate of 8.55 kbps, selectable mode vocoder (SMV) (2001) and EVRC-B (2005) at a peak rate of 8.55 kbps with multiple operational modes to tradeoff between capacity and voice quality. The 3GPP2 WB family of codecs consists of multi-rate wideband (VMR-WB) (2004) at a peak data rate of 13.3 or 8.55kbps and enhanced variable rate vocoder wideband extension (EVRC-WB) (2006) [5] at a peak data rate of 8.55 kbps.

5. THE NEXT KILLER APPLICATION

The marriage of voice over internet protocol (VoIP) and wireless technology has created a lot of attention over the past few years. Such methodology will provide convergence which is the key technology push or “the next killer application” in migrating today’s separate circuit and packet switched voice networks to a unified core network. The technology has just started to make an impact in the telecommunication as a whole and most of the industry predicts that it is going to revolutionize the whole telecommunication market. Although there are numerous challenges that remain to be solved, the infrastructure and core network appear to be shaping up. Internet multimedia subsystem (IMS) [1] and multimedia domain (MMD) [3] address some of these challenges by providing common components and interoperability between networks independent of their access points. IMS framework supports standard session initiation protocol (SIP) that can provide multi-vendor networks with well-defined interfaces between different terminals. Entities such as voice call continuity (VCC) make it possible to offer transparent voice services for a multi-radio hierarchical network including wireless local area network (WLAN) and worldwide interoperability for microwave access (WiMAX). IMS addresses QoS issues such as bandwidth fluctuation, congestion, heavy loss of packets, varying jitter, handoffs and even security, but more remains to be solved. IMS also enables a lot of other multimedia services such as push to talk and video share by working at the higher level of the protocol stack.

6. CONCLUSIONS

Today, voice services are still predominantly provided by narrowband circuit switched technology in both the fixed and mobile domains. Although we are witnessing VoIP services with wideband voice quality gaining grounds within the fixed operators’ domain but we are yet to see a similar drive among the wireless operators. IMS framework will obviously take some time to complete, but when it does, it will provide a cost effective solution to the operators. They can leverage their open architecture by adding enhanced and new software-based services without the disruption and expense of replacing any equipment. FMC provides quality, reliability, scalability and flexibility. It offers quality by giving an opportunity to deploy unified wideband speech codec across all networks without transcoding and with voice quality that is much better than the traditional toll quality. It offers reliability by providing a transparent voice service which will be consistent with the traditional telephone network. It offers scalability by removing the dependencies on media gateways for transcoding. It offers flexibility by allowing solutions that support features beyond traditional PCM voice with new capabilities such as conferencing, presence and simultaneous voice with data and multimedia services.

7. REFERENCES

[1] http://www.3gpp.org/; 3GPP TS 26.103 V6.2.0 (2006-03) Speech codec list for GSM and UMTS.

[2] B. Bessette et al., "The Adaptive Multirate Wideband Speech Codec (AMR-WB)," IEEE Trans. Speech and Audio Processing, vol. 10, no. 8, Nov 2002, pp. 620–36.

[3] http://www.3gpp2.org/

[4] Comparing CDMA2000 and GSM/GPRS/EDGE/UMTS, CDMA development group Dec. 2005. http://www.cdg.org

[5] V. Krishnan et al., 'EVRC-Wideband: The New 3GPP2 Wideband Vocoder Standard', ICASSP 2007.

[6] http://www.ietf.org/

[7] http://www.itu.int/ITU-T/ ITU-T G.729.1 Recommendation, May 2006.

[8] S. Ragot et al., “ITU-T G.729.1: An 8-32 kbit/s Scalable Coder Interoperable with G.729 for Wideband Telephony and Voice Over IP,” ICASSP 2007.