Tandem Voice Communications: Digital Cellular, VoIP, and Voice over Wi-Fi*

Jerry D. Gibson
Electrical & Computer Engineering
University of California, Santa Barbara
Santa Barbara, CA 93106-6065

Bo Wei
Department of Electrical Engineering
Southern Methodist University
Dallas, TX 75275

***Abstract*- We consider the problems of voice over wireless LANs and voice communications over heterogeneous asynchronous tandem networks, including digital cellular, VoIP, and voice over Wi-Fi. For voice over Wi-Fi, we minimize retransmissions by combining new packetization methods and packet loss concealment approaches. We demonstrate that tandem network connections can suffer significant loss in voice quality, even with ideal channels. We present a particular example of a VoIP network in tandem with voice over Wi-Fi and study end-to-end performance for different voice codecs, bit error rates, packet loss rates, and packet loss concealment methods.**

I. INTRODUCTION

Voice is the method of choice for real time communications [1]. Voice is so important to human communications that we have constructed entire networks centered around voice, namely, the public switched telephone network (PSTN) [2] and the analog/digital cellular networks [3]. Computer networks were originally developed with data transmission in mind, but now there is burgeoning interest in transmitting various other multimedia services over the Internet, including voice [5]. Indeed, voice over the Internet Protocol (VoIP) is growing rapidly and is expected to do so for the near future. A new and powerful development for data communications is the emergence of wireless local area networks (WLANs) in the embodiment of the 802.11 a, b, g standards [6, 7], collectively referred to as Wi-Fi [7]. Because of the proliferation and expected expansion of Wi-Fi networks, considerable attention is now being turned to voice over Wi-Fi, with some companies already offering proprietary networks, handsets, and solutions [8-10].

Based on this discussion, one can envision that future voice communications might consist of a digital cellular user communicating with a voice over Wi-Fi user through a VoIP backbone, resulting in the tandem connection of a digital cellular network into a VoIP backbone followed by a voice over Wi-Fi link, as shown in Fig. 1, and the reverse connection as well. Since each of these networks may use different voice codecs and different network protocols, as well as have different channel behaviors and bit error/packet loss mechanisms, it is not evident that acceptable, bandwidth efficient voice service is possible over such a connection.

*This research was supported by the National Science Foundation under Grant Nos. CCR-0243332 and CCR-0087568, and by the California MICRO Program, Dolby Laboratories, Inc., Lucent Technologies, Inc., Mindspeed Technologies, Inc., and Qualcomm, Inc.

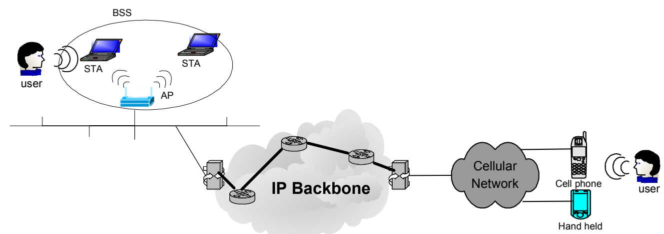

Figure 1. Voice over Wi-Fi, VoIP, and Digital Cellular Tandem Connection

In fact, even if the channels involved are near-ideal, one is still faced with the well-known problem of asynchronous tandem connections of voice coders [3, 11]. Additionally, although the digital cellular networks utilize relatively sophisticated error detection and correction, unequal error protection, and error concealment, the error control, packet loss mechanisms, and packet loss concealment methods for VoIP and voice over Wi-Fi are relatively primitive. Furthermore, while there is some work in progress on protocol design and priority-based services for the individual Wi-Fi links, very few studies have addressed the performance and design of such cascaded network connections for real-time voice communications, which appear to be the end-to-end networks of our future.

In Sec. II, we discuss the voice codecs, bit error/packet loss characteristics, and concealment methods for the PSTN, digital cellular, and VoIP systems. We then develop the key issues in voice over Wi-Fi in Sec. III, followed by a particular example of VoIP connected in tandem with a voice over Wi-Fi system. Conclusions are presented in Sec. V.

II. NETWORKS FOR VOICE COMMUNICATIONS

In this section, we describe the relevant details of current and developing networks for voice communications with respect to voice codecs, channels, protocols, bit error or packet loss mechanisms, and key design parameters.

A. The Public Switched Telephone Network (PSTN)

The voice codec most often used in the PSTN is 64 kilobits/sec. (kbits/s) logarithmic pulse code modulation (log-PCM), designated by the ITU-T as G.711, and which is taken as the standard for toll quality voice transmission. The time division multiplexed (TDM) links in the PSTN are very reliable with bit error rates (BERs) of 10^{-6} to 10^{-9}. Furthermore, G.711 is designed with several asynchronous tandems in mind, and even 8 asynchronous tandems of G.711 with itself has been shown to still maintain a Mean

Opinion Score greater than 4.0 when a single encoding is 4.4 to 4.5 [12].

Other voice codecs have been standardized for the PSTN over the years. These include G.721 (now G.726), G.727, G.728, G.729, and G.729A [12, 13] for narrowband (telephone bandwidth) speech (200 to 3400 Hz) and G.722, G.722.1 [14], and G.722.2 [15] for wideband speech (50 Hz to 7 kHz). Table I presents some results concerning the tandem performance of the narrowband speech codecs. Note that asynchronous tandem connections of these codecs with themselves do not cause an unacceptable loss in performance compared to a single encoding, although the MOS for 4 tandems of G.726 and 3 tandems of G.729 drops considerably.

Table I. Representative Asynchronous Tandem Performance of Selected PSTN Codecs

Voice Codec	Mean Opinion Score (MOS)
G.711x4	>3.50
G.726x4	2.91
G.729x2	3.27
G.729x3	2.68

B. *Digital Cellular Networks*

Digital cellular networks provide wireless voice connectivity by combining high quality voice codecs, unequal forward error correction of sensitive bits, error detection and concealment of uncorrected errors, and interleaving. Table II shows available results for multiple tandem encodings of the codecs, including results from tandem connections of these codecs with the PSTN backbone codecs in Table I. It is clear that tandem encodings result in a drop in performance as seen in the lower MOS values. Furthermore, tandem encodings add to the end-to-end delay because of the algorithmic delays in decoding and re-encoding.

Tandem encodings are not discussed often within digital cellular applications since the codec for the backbone wireline network is often assumed to be G.711. However, it is recognized that tandem encodings with codecs other than G.711 can lead to a loss in performance and that tandem encodings constitute a significant problem for end-to-end voice quality [3, p. 444; 11]. In particular, transcoding at network interfaces and source coding distortion accumulation due to repeated coding has been investigated with the goal of obtaining a transparent transition between certain speech codecs [16-18]. Some system-wide approaches also have been developed [19] for specific networks. The general tandeming/transcoding problem remains open.

The melding of voice coder design, forward error correction and detection, unequal error protection, and error concealment in digital cellular has important lessons for designing voice communications systems for VoIP and voice over Wi-Fi.

Table II.Representative Asynchronous Tandem Performance of Selected Digital Cellular Codecs

Voice Codec	Mean Opinion Score (MOS)
IS-641x2	3.62
GSM-EFRx2	4.13
IS-641+G.729	3.48
GSM-FR+G.729	3.05
GSM-EFR+G.729	3.53
GSM-EFR+G.729+G.729	3.21
IS-641+G.729+G.729	3.10

C. *Voice Over Internet Protocol (VoIP)*

Among the issues in developing good VoIP systems are voice quality, latency, jitter, packet loss performance, packetization, and the design of the network [5]. Broadly speaking, the voice codec in VoIP systems should achieve toll or near toll quality, have as low a delay as possible, and have good resilience to packet loss. ITU-T Recommendation G.114 provides specifications for delay when echo is properly controlled [20]. Interestingly, voice codecs used in prominent VoIP products are all ported from previous standards and other applications. Today's VoIP product offerings typically include G.711, G.729, and G.722, in addition to G.723.1. See Table III for a summary of the relevant properties offered by each coder.

Table III. Properties of Common VoIP Codecs

Codec	Relevant Properties
G.711	Low delay, toll quality, low complexity, higher rate
G.729	Toll quality, acceptable delay, low rate, acceptable complexity
G.723.1	Low rate, acceptable quality, relatively high delay
G.722	Wideband speech, low delay, low complexity, higher rate

The coders in Table III, as a set, offer a full range of alternatives in terms of rate, voice quality, complexity, and delay. What is not evident from this table is how effectively one can conceal packet losses with each of these coders. Packet loss concealment is particularly important since in order to reduce latency, retransmissions are not allowed in VoIP.

Rather recently, a packet loss concealment algorithm has been developed for G.711 [21]. Based upon 10 ms packets and assuming the previous frame was received correctly, the method generates a synthesized or concealment waveform from the last pitch period with no attenuation. If the next packet is lost as well, the method uses multiple pitch periods with a linear attenuation at a rate of 20% per 10 ms. After 60 ms, the synthesized signal is zero.

G.729 and G.723.1 suffer from the problem that the predictor parameters (line spectral frequencies) are predictively encoded from frame-to-frame. For G.729, the basic approach to packet loss concealment if a single 10 ms frame is erased is: (i) generate a replacement excitation based upon the classification of the previous frame as voiced or unvoiced, (ii) repeat the synthesis filter parameters from the previous frame, and (iii) attenuate the memory of the gain predictor.

III. Voice Over Wi-fi

Wireless local area networks (WLANs), such as 802.11b, 802.11a, and 802.11g, are becoming extremely popular for use in businesses, homes, and public places. As a result, there is considerable interest in developing VoIP for Wi-Fi, which we designate here as voice over Wi-Fi. The issues involved for voice over Wi-Fi are much the same as for VoIP including voice quality, latency, jitter, packet loss performance, and packetization, and all play a critical role. However, the physical link in Wi-Fi is wireless, and as a result, bit errors will commonly occur and this, in turn, effects link protocol design and packet loss concealment.

One issue, in particular, is how to handle packets with bit errors. One approach would be to detect bit errors in a packet using a cyclic redundancy check (CRC) error detection code [4], and request a retransmission if a bit error is detected in the packet. In fact, the IEEE 802.11 MAC layer defines two different access methods. One method, called the distributed coordination function (DCF), is basically carrier sense multiple access with collision avoidance (CSMA/CA), and the basic access scheme is, if the channel is sensed to be idle, the node starts its transmissions. A CRC is computed over the entire received packet and an acknowledgment is sent if the packet is error-free. If an error is detected in the packet, a retransmission is requested. Up to 7 retransmissions for short packets and up to 4 retransmissions for large packets are allowed. This method clearly adds to latency and is in contrast to avoiding retransmissions altogether in VoIP, but how does one deal with bit errors? The answer lies in a combination of those techniques used in digital cellular in conjunction with different packetization and packet loss concealment methods.

An efficient way to transmit voice over WLANs is to employ a reservation scheme that guarantees delay and bandwidth. Work is underway on a new standard, 802.11e, which is designed to support delay-sensitive applications with multiple managed levels of quality of service (QoS) for data, voice, and video. However, the focus of this standard on priority-based QoS does not address the choice or design of the voice codec or the packet loss concealment issues for voice over Wi-Fi.

Several recent research efforts have targeted improving voice communications over WLANs and over mobile ad hoc networks by involving the codec choice and packet loss concealment mechanisms. Servetti and De Martin [22] adapt the rate of the speech codec, the narrowband AMR coders, based upon the estimated instantaneous channel conditions, with the goal of using a higher rate and larger packets when the channel is good and a lower rate and shorter packets when the channel is poor. The no retransmission and two retransmission cases are considered and header compression is used to reduce the packet overhead. In comparison to a constant bit rate approach, their method reduces packet losses and delays by reducing retransmissions.

In [24], the G.729 codec is used, and based upon the packet loss concealment method employed in G.729, it is shown that the location of lost speech frames in terms of the speech mode is important to how long it takes for the decoder output to recover. To address these issues, the authors use the voicing decisions from G.729 and consider a higher retry limit on sensitive packets and duplicate transmission of sensitive packets. Both schemes provide improvements, with the combination of the two yielding the best performance.

A recent effort by Petracca, Servetti, and De Martin [23] uses analysis-by-synthesis to develop an estimate of the importance of each packet of speech compressed by the NB-AMR codec at 12.2 kbits/s. If the packet is determined to be perceptually important, it is transmitted more than once in order to provide an increased likelihood that the packet will be successfully received. The standard NB-AMR packet loss concealment method is employed. Repetition of the perceptually most important 10% of the packets is shown to improve voice quality.

In [25], selective error checking of bits produced by the NB-AMR codec at 7.95 kbits/s is used to modify the MAC layer retransmission scheme for multiple hop wireless networks without intervening source reconstruction. In particular, the NB-AMR codec classifies bits to be transmitted into two classes, Class A and Class B, according to their sensitivity in terms of reconstructed speech quality. In [25], a modified MAC CRC checks the header and the Class A bits for errors. If an error is detected in these bits, a retransmission is requested. The Class B bits are not subject to a CRC. It is shown that this method can reduce the packet loss rate for multiple wireless hops and that the speech produced with the modified protocol is preferred by listeners in informal listening tests.

A system is proposed in [28] that uses the NB-AMR and a recursive systematic convolutional code together with a packetization procedure the spreads encoded bits over N packets. Improved performance is obtained compared to G.711, but a delay of N packets is incurred.

IV. A Tandem Connection Of VoIP And Voice Over Wi-fi

Here we consider the tandem connection of a wireline VoIP network with a voice over Wi-Fi link and examine some of the trade-offs among the choice of voice codecs, the packet loss concealment (PLC) method, and the interaction with the MAC layer protocol. We focus on the VoIP into voice over Wi-Fi direction. In general, the operations are not symmetric with respect to direction of transmission. Two cases are considered for the choice of speech codec in the wireline VoIP connection, G.711 and G.729, both of which are widely available in offered services and products.

To begin, we consider the case where the speech codec in the backbone network is G.711. Among the advantages for G.711 are packetization flexibility, payload efficiency, and speech quality in VoIP [26]. The RTP/UDP/IP protocol is used for the wireline network and packet loss occurs when packets arrive later than some delay constraint. No retransmissions are allowed and packet losses are concealed using the method in G.711, Appendix I [21].

As a benchmark system, we utilize G.711 in the WLAN as well. Each 10 ms frame (80 bytes) is packetized into a WLAN packet, and the receiver checks the header CRC and retransmits if the CRC fails. Otherwise, the CRC for the speech frame is checked, and the entire packet is dropped when an error is detected in the speech data--no retransmissions are requested for a CRC failure over the speech data. The PLC procedure in [21] is applied to reconstruct the packet at the output of the WLAN link.

We measure the end-to-end speech quality when the packet loss rate (PLR) in the wireline link and the bit error rate (BER) in the wireless link vary. Note that bit errors in the WLAN link can translate into packet losses as described earlier. The Enhanced Modified Bark Spectral Distortion (EMBSD) objective speech quality metric [27] is applied to generate MOS values by comparing the coded and processed speech with the original. The simulation results for the system where G.711 is used on both links, are shown in Fig. 2. Note that higher values of MOS indicate better performance. Clearly, when both channels are good, the output speech quality is excellent. This is expected since the only distortion mechanism in this situation is G.711 connected in an asynchronous tandem, and that is known to perform well. As the bit error rate in the wireless link increases, however, the tandem performance drops precipitously.

We now turn to the case where in the WLAN, we use the NB-AMR voice codec at the 5.9 kbits/s and 12.2 kbits/s rates and switch between them based upon channel conditions. Coded bits are divided into two groups, the most subjectively significant Class A bits and less significant Class B bits. Unequal error protection (UEP) and selective error checking are used by constructing a modified speech packet for the WLAN as shown in Fig. 3.

The packet dropping procedure in the WLAN is: The CRC in the MAC header only checks the header section instead of the entire packet. If this CRC check fails, the packet is dropped and retransmission is required. The packet is kept otherwise. For the kept packets, after the channel decoding process for Class A bits, we go through the CRC check for Class A bits. If this CRC fails, error concealment is applied; otherwise, source decoding is performed even if there are errors in Class B bits. This method reduces retransmissions since only a small portion of a packet is subject to the CRC for a retransmission request.

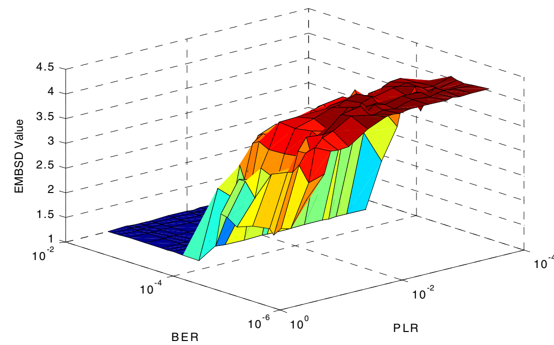

Figure 2. G.711 VoIP into G.711 Voice Over Wi-Fi Performance

RTP/UDP/IP/MAC Header	FEC protected (Class A bits+CRC)	Class B bits

Figure 3. WLAN Packet Format for NB-AMR Codec

Class A bits in the AMR 12.2 kbits/s codec are not protected by any FEC coding, but a rate ½ convolutional code is applied to the 5.9 kbits/s mode. This results in a 32 byte payload for the 12.2 kbits/s mode and a 20 byte payload for the 5.9 kbits/s mode. To focus on the tradeoff between source and channel distortion over the WLAN, we de-emphasize the delay constraint by assuming all packets pass the header CRC check within a certain delay period (50 ms for instance) through retransmissions. Error concealment for the AMR codecs follows the AMR channel coding standard.

From Fig. 4, we can see that if we implement the adaptation between the two NB-AMR modes, better performance in terms of the MOS value over a wider range of channel conditions is obtained compared to the system that uses G.711 in the WLAN, and also in comparison to using only one of the two AMR codecs. The 12.2 kbits/s mode performs well for better channels, although slightly poorer than the G.711 into G.711 tandem, but starts to perform poorly for a BER over the WLAN around 10^{-4}. It is then preferable to switch to the lower rate voice codec, thus freeing up bits to be used for error control coding.

The performance with G.711 in the backbone and the NB-AMR codecs in the WLAN can be contrasted with the case where G.729 in used in the wireline VoIP system. As usual, the RTP/UDP/IP protocol is used without retransmissions for the wireline network and packet loss occurs when packets arrive later than some delay constraint. The standard G.729 packet loss concealment method is applied. The results are shown in Fig. 5. Comparing Figs. 4 and 5, what is immediately evident is that the overall voice quality is significantly lower, primarily due to the tandem connection of the G.729 codec with the NB-AMR codecs at 12.2 kbits/s and 5.9 kbits/s.

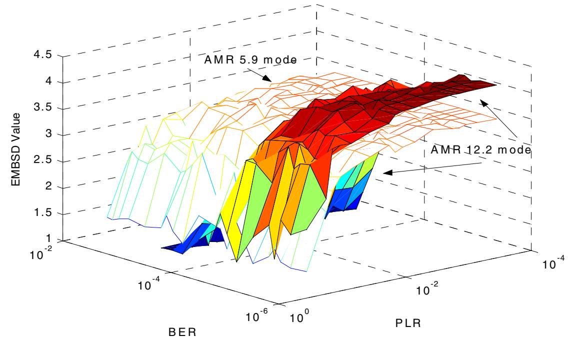

Figure 4. Adaptation Between AMR 12.2 mode and AMR 5.9 Codecs in a WLAN in Tandem with G.711 VoIP

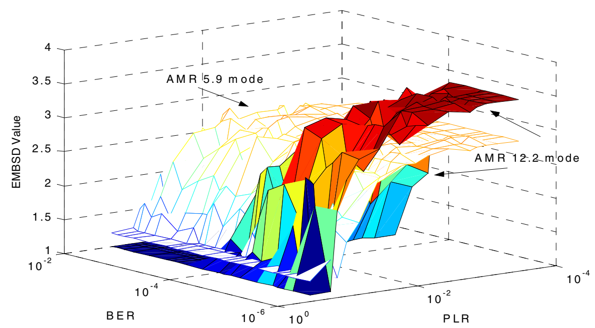

Figure 5. Performance with G.729 in the Wireline VoIP and the NB-AMR 5.9 and 12.2 kbit/s Codecs in the WLAN

V. CONCLUSIONS

We have presented an approach to packet loss concealment in voice over Wi-Fi that minimizes retransmissions and provides good voice quality over a wide range of channel conditions. Additionally, we have examined the tandem connection of a VoIP network with voice over Wi-Fi and highlighted the issues involved in maintaining acceptable end-to-end voice communications. It is evident that if a digital cellular link is included in tandem with the VoIP and voice over Wi-Fi networks, an additional loss in performance can be expected.

REFERENCES

[1] B. Teitelbaum, "Leading-edge voice communications for the MITC," Sept. 12, 2003 at http://people.internet2.edu/~ben/.

[2] J. C. Bellamy, *Digital Telephony*, John Wiley & Sons, 2000.

[3] T. S. Rappaport, *Wireless Communications: Principles and Practice*, Prentice-Hall, second edition, 2002.

[4] L. L. Peterson and B. S. Davie, *Computer Networks: A Systems Approach*, Morgan Kaufmann, third edtion, 2003.

[5] B. Goode, "Voice Over Internet Protocol (VoIP)," *Proceedings of the IEEE,* vol. 90, pp. 1495-1517, Sept. 2002.

[6] ISO/IEC and IEEE Draft International Standards,"Part 11: Wireless LAN Medium Access Control (MAC) an dPhysical Layer (PHY) Specifications," ISO/IEC 8802-11, IEEE P802.11/D10, Jan. 1999.

[7] http://wi-fiplanet.webopedia.com/TERM/w/Wi_Fi.html

[8] http://www.cisco.com

[9] http://www.spectralink.com

[10] http://www.globalipsound.com

[11] S. F. Campos Neto and F. L. Corcoran, "Performance assessment of tandem connection of enhanced cellular coders," *Proc. ICASSP*, Phoenix, AZ, 1999,

[12] W. R. Daumer, "Subjective evaluation of several efficient speech coders," *IEEE Trans. on Communications,* vol. COM-30, pp. 655-662, April 1982.

[13] A. S. Spanias, "Speech coding standards," Chap. 3 in *Multimedia Communications:Directions and Innovations,* J. D. Gibson, ed., Academic Press, 2001, pp. 25-44.

[14] ITU, "Coding at 24 and 32 Kbit/s for hands-free operation in systems with low frame loss", Sept.1999.

[15] ITU-T Recommendation G.722.2 (2002) Wideband coding of speech at around 16 kbit/s using Adaptive Multi-Rate Wideband (AMR-WB).

[16] T. Nishitani, "Tandem transcoding without distortion accumulation," *IEEE Trans. on Commun.*, vol. COM-34, pp. 278-284, Mar. 1986.

[17] S. M. Tsai and J. F. Yang, "GSM to G.729 speech transcoder", *Proc. IEEE Intl. conference on Electronics, Circuits and Systems*, Vol. 1, pp. 2-5, 2001.

[18] H. G. Kang, H. K. Kim and R. V. Cox, "Improving the transcoding capability of speech coders", *IEEE trans. on Multimedia*, pp. 24-33, March 2003.

[19] *ETSI*, "Digital Cellular Telecommunications System (Phase2+); Inband Tandem Free Operation (TFO) of Speech Codecs;(GSM 08.62 version 7.0.0 Release 1998)," 1998.

[20] ITU-T, One-Way Transmission Time,Recommendation G.114, May 2000.

[21] ITU-T, G.711, Appendix I: A high quality low-complexity algorithm for packet loss concealment with G.711," Sept. 1999.

[22] A. Servetti and J. C. De Martin, "Adaptive interactive speech transmission over 802.11 wireless LANs," *Proc. IEEE Int. Workshop on DSP in Mobile and Vehicular Systems,* Nagoya, Japan, April 2003.

[23] M. Petracca, A. Servetti, and J. C. De Martin, "Voice transmission over 802.11 wireless networks using analysis-by-synthesis packet classification," *First Int. Symp. On Control, Communications, and Signal Processing,* Hammamet, Tunisia, March 21-24, 2004.

[24] C. Hoene, I. Carreras, and A. Wolisz, "Voice over IP: Improving the quality over wireless LAN by adopting a booster mechanism - an experimental approach," *ITCOM,* 2001.

[25] H. Dong, I. D. Chakares, A. Gersho, E. Belding-Royer, and J. D. Gibson, "Selective bit-error checking at the MAC layer for voice over mobile ad hoc networks with IEEE 802.11," *IEEE Wireless Communications and Networking Conference*, Atlanta, GA, 2004.

[26] J. Hagenauer and T. Stockhammer, "Channel Coding and Transmission Aspects for Wireless Multimedia," *Proceedings of IEEE*, Vol. 87, No. 10, October 1999.

[27] W. Yang, M. Benbouchta, and R. Yantorno, "Performance of the modified bark spectral distortion as an objective speech quality measure," *Proc. ICASSP*, Seattle, 1998, pp. 541-544.

[28] M. Kaindl and N. Gortz "AMR Voice Transmission over Mobile Internet", *Proceedings ICASSP 2002*, vol. II, pp. 2049 – 2052, May 2002.