Real-Time speech with AX.25 -an evaluation of the possibilities and restrictions

Nicolai Heilemann, DL9TS

{DL9TS@QSL.NET}

Reutlingen University, Department of Electronics

Amateur Radio Laboratory

July, 2003

A study about the possibilities of real-time speech transmission using amateur packet radio is presented. In this study it is shown, that a common 9600 Bit/s- channel has the theoretical possibility of transmitting speech data with only short delays. The restrictions of the AX.25 protocol for real-time transmission are discussed, showing that an appropriate setting of the transmission parameters still allows real-time transmission. A hybrid emulation system using traces from a real wireless channel as well as coded speech gives an estimation of the expected speech quality and an implementation of a real-time transmission tool allows first experiments. Preliminary results of these experiments show the possibility to use AX.25 for real-time speech transmission even if the resulting speech quality is not yet satisfying. However the tolerance of listeners towards distortions has been much higher than expected and there are possibilities to improve the system with respect to the user perceived speech quality.

Key words Real-Time Speech Transmission, AX.25, Speech Quality

1 Introduction

In the last few years, real-time speech transmission over packet based networks (especially Voice over IP - VoIP) has become a major topic. Much research is performed tackling various aspects of this technology. Especially wireless use of packet based speech transmission has become a novel technique with lots of possibilities but also restrictions.

Being able to experiment with their own packet based wireless network, radio amateurs can gain valuable experience and so can contribute interesting results to the development of this technique. Still the amateur radio packet based network is heavily restricted, compared to commercial nets (eg. Ethernet-based systems). Especially the available bandwidth of this network is very low so that the demands for the used application are very strict. This does not have to be a drawback. Using a very restricted system allows to test the boundaries of a technology under the strictest conditions which might never occur in systems with larger resources, thus allowing to optimise the parameters.

This study aims to show that the packet radio network is able to be used for experiments with

real-time voice communication. As a first step the restrictions of a wireless channel and their impact on speech have to be evaluated. The effects of routing and multiple hops as well as multiplexing of real-time speech with concurrent traffic are left aside at this stage of the work. For the experiments a 9.6 kBit/s AX.25 [BNT97] connection with standard packet radio equipment for 430MHz has been chosen (T7F transceiver [Eck99] and YAM modem [Pal99]). The transmission distance was approximately 3 km with no direct sight.

The remainder of this paper is organised as follows: Section 2 gives some general considerations about the problems which are expected to occur with the chosen set-up. Section 3 presents an emulation which gives first estimations whether a real-time speech transmission system will ever work with a 9.6 kBit/s wireless channel. Section 4 presents the used implementation of a Voice over AX.25 (VoAX.25) software based on a common audio tool. This leads to section 5 which presents the results of first experiments with the system. Section 6 concludes the paper by offering ideas to overcome the problems found in the study.

2 General considerations

Modern codecs easily allow to compress speech data to very low data rates. For example the chosen linear predictive codec can generate a data stream of 5.6 kBit/s and there are ACELP (advanced codebook excited linear predictive) codecs with even lower data rates. So the transmission with standard AX.25 equipment sending with 9.6 kBit/s does not appear to be a major problem.

However there are different side effects of packet based speech transmission which have to be considered:

1. Packet header overhead:
   Being a packet based routing protocol, AX.25 requires control information to be contained in each packet that is sent through the network.

In AX.25 this control information uses plain text (amateur radio call signs) to encode the sender's, repeater's and destination's address. Therefore these address fields take a lot of space and the AX.25 frame can contain up to 33 octets of control information.

In order to guarantee low delays, confirmed submission (which is normally used in AX.25) cannot be accepted. On the contrary a data stream of "UI"("unnumbered information")-frames has to be used (cf. below). The name "unnumbered information" indicates, that there is no sequence number contained in the header, so an additional header has to be added for flow control. The most common protocol for flow control of real-time sessions is the Real-Time Transfer Protocol (RTP) [SCFJ96] which adds 12 octets of control information to each frame.

Besides the speech information, some additional information (e.g. reception reports or additional information about the sender) has to be transmitted using the Real-Time Control Protocol (RTCP), a protocol closely linked to RTP. This control information is transmitted in UI-frames similar to the ones used for the speech data. In order to separate speech from control data an additional header of one octet is used (cf. section 4).

Adding up all the header information shows, that the size of the header dominates the AX.25 frame compared to the speech payload. Eg. the chosen codec (linear predictive codec) uses 14 octets per frame containing 22.5ms of speech and a header of 46 octets is needed per packet. Thus the header must not be neglected when estimating whether it is possible to transmit real-time speech with standard AX.25 equipment:

- AX.25 frame: $p_a = 33 * 8\,Bit$
- RTP header: $p_r = 12 * 8\,Bit$
- Additional control header: $p_c = 1 * 8\,Bit$
- Speech payload:
  $p_p = 14 * 8\,Bit$, $t_p = 22.5\,ms$

Sending each speech frame in its own AX.25 packet leads to a packet size of

$$p_1 = p_a + p_r + p_c + p_p = 480\,Bit \qquad (1)$$

This packet is generated every $t_p = 22.5\,ms$, leading to a bitrate of

$$b_1 = \frac{p_1}{t_p} = 21.3\,\frac{kBit}{s} \qquad (2)$$

This is obviously much too much even for a 19.2 kBit/s transmission channel.

The sole solution to this problem is transmitting more than one speech frame per AX.25 packet which reduces the packet header overhead.

For example 4 speech frames per packet result in

$$p_4 = p_a + p_r + p_c + 4 * p_p = 816\,Bit \qquad (3)$$

This packet is generated only every

$$4 * t_p = 90\,ms$$

which leads to an effective bitrate of

$$b_4 = \frac{p_4}{4 * t_p} = 9.066\,\frac{kBit}{s} \qquad (4)$$

This solution causes a larger delay. Now the speech is not transmitted directly, but the sender has to collect four speech frames before sending them. Still a delay of 90ms can still be considered as being real-time conversation, thus this disadvantage can be accepted here.

2. User perceived packet loss

Speech codecs based on vocal tract models like the used linear predictive codec or modern ACELP codecs highly compress the speech in order to achieve low transmission rates. This leads to an absence of any redundancy within the speech data, and it is very important for the listener's perception that no data is lost. However in a wireless network like packet radio one cannot assume the interlinks between two nodes to be loss-less. Bit errors due to natural or technical interference will always occur and will force the affected packet to be dropped at the receiver. In this study only a point-to-point transmission is used. Therefore no interference with concurrent traffic has to be taken into consideration at this stage of the work. However for practical use an empty channel must not be assumed. Other traffic streams using the same transmission channel will cause packet collisions and further packets to be lost.

Normally lost packets will not affect packet radio communication too heavily as long as their percentage remains small. The receiver detects the packet loss and reorders the lost parts. Yet this is not possible in a real-time transmission. The confirmed transmission of the balanced mode of AX.25 leads to much too large delays for speech communication. Therefore UI-frames are used here, which do not require confirmation and are sent without waiting for a "receive ready" from the receiver. This means, that neither the sender nor the receiver has control whether the packet will reach its destination. If a packet is damaged or routing fails, the packet is lost. The receiver can detect the fact of packet loss by looking at the sequence number of the RTP header, still reordering of the missing parts is useless as the speech has to be played out to the listener immediately, so a reordered part would reach its destination much too late.

Different studies have shown the impact of packet loss on the user perceived speech quality [SWLI01, SLHM01]. As large packets (i.e. large number of speech frames per AX.25 frame) have to be used in order to reduce the packet header overhead, one lost AX.25 frame will result in a significant gap in the speech (eg. 90ms). This cannot remain unrecognised and will lead to an unsatisfying speech quality.

There are techniques to reduce the impact of packet loss. Sender based loss protection (eg.

FEC - forward error correction) works by adding extra redundant information to the transmission, which is not possible here due to bandwidth limitations. Receiver based loss protection restores lost parts from received data. Yet [SLHM01, SWLI01] have shown, that if several speech frames are lost in a bulk, this loss protection system has increasing problems to conceal the perception of the loss. Therefore the loss concealment algorithm implemented into the audio tool can be expected to deliver only a low performance as large packets are used here and the loss of one packet leads to the loss of various speech frames.

3 Emulation of AX.25 speech transmission

In order to verify the considerations presented above, a hybrid simulation system which allows to combine real speech with various kinds of network channel simulators has been set up. This system can be used as an emulator, which applies the characteristics determined from a real transmission channel directly to speech. Thus speech samples are produced which contain exactly the errors of this channel. Figure 1 gives an overview of the structure of the emulation system:

- The AX.25 transmission channel is evaluated for its loss characteristics using an AX.25 traffic generator and an appropriate monitor program. The settings of the traffic generator are chosen to produce packets of the same packet size and transmission interval as packets containing coded speech (eg. packet size = 816 Bit, packets transmitted every 90ms). This results in a trace file containing the errors which occurred during the transmission.
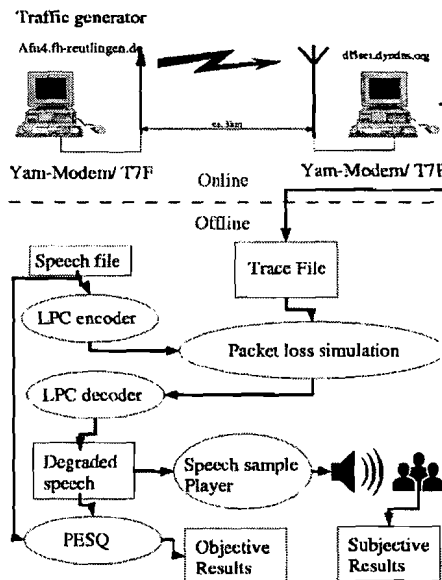

Figure 1: Overview of the emulation system

- Speech samples are coded using a LPC simulation coder producing a file representing the bit stream which has to be transmitted to the receiver.

- A channel simulator which is controlled by the trace file applies the errors to the coded speech. Therefore a bit stream with distortions is produced which is an exact representation of the speech which would have been sent through the wireless channel instead of the test packets.

- Finally the bit stream file is decoded.

The distorted speech can be evaluated by various means. It can be listened to directly for subjective listening tests using the well-known MOS (mean opinion score)-scale [ITU96] or objective computational methods like PESQ (perceptual evaluation of speech quality) [RBHH00] can be used.

Table 1 presents the emulation results for a packet size of 4 speech frames. There are some significant

Table 1: Results of the emulation for
4 speech frames/AX.25 packet

min. packet loss	0%
max. packet loss	54%
av. packet loss	13.6%
max. PESQ MOS	2.825
min. PESQ MOS	1.751
av. MOS	2.345

points to observe: The most obvious result is the wide range of the loss percentage which occurred. It is easily imaginable that no communication will be possible when more than 50% of the speech data is lost. The results show that these cases occur and have to be considered. Still the loss percentage is much lower most of the time which is indicated by the average packet loss rate. So one can hope that communication is possible even if more than 10% loss will result in heavily distorted speech and thus a poor communication quality.

These results match with the evaluation of the speech quality using PESQ: There were cases in which the speech quality was marked as insufficient (MOS < 2 indicates a completely incomprehensible speech) but in average the score indicates that the speech is intelligible, even if the speech quality is at most fair. The maximal MOS of only 2.825 seems low, yet this limit is caused by the used codec. Linear predictive coding techniques only allow synthetic speech quality with MOS below 3. Choosing a different codec here (eg. an ACELP codec) will increase the speech quality in general.

4 Speech transmission system

After proving the real-time abilities of an AX.25 channel by emulation, an AX.25 speech transmission system has been set up for first practical tests. A modular implementation has been chosen, keeping the audio tool separated from the AX.25 network

driver. This allows to replace the different parts of the implementation independently.

A common audio tool for internet telephony (the Robust Audio Tool -RAT- by University College London [rat02])has been chosen, which communicates with the AX.25 network over an IP to AX.25 converter as shown in Figure 2.

This solution seems to be a rather difficult, yet is has various advantages compared to a proprietary Voice over AX.25 tool:

- The implementation is almost independent from the used audio tool. Therefore tools on different platforms can be used as long as they support the standard real-time protocols for IP (i.e. RTP/RTCP).

- There is a larger number of available codecs for various experiments, as one can use the codecs implemented in the chosen audio tool and does not have to implement each of them separately. If different audio tools support different codecs one can exchange them quickly without changing the AX.25 protocol implementation.

- The tool works independently from AX.25 (i.e the AX.25 protocol stack is handled externally) therefore it can still be used for Voice over IP.

- As the audio tool and the AX.25 network adaptor communicate by using IP, they do not have to be located on the same host.

The comparable inconvenient process of first packing the speech into IP datagrams, unpacking it again and finally sending it as AX.25 frames adds an extra delay which is difficult to determine. Still [CS00] has shown, that processing of UDP and IP packets in the TCP/IP stack adds delays of several μs. So the extra delay of this process is far below 1% compared to the packetisation delay of 90ms and is neglected here.

In order to exchange additional information, RTP based tools use the Real-Time Control Protocol (RTCP). This protocol allows the receiver to send
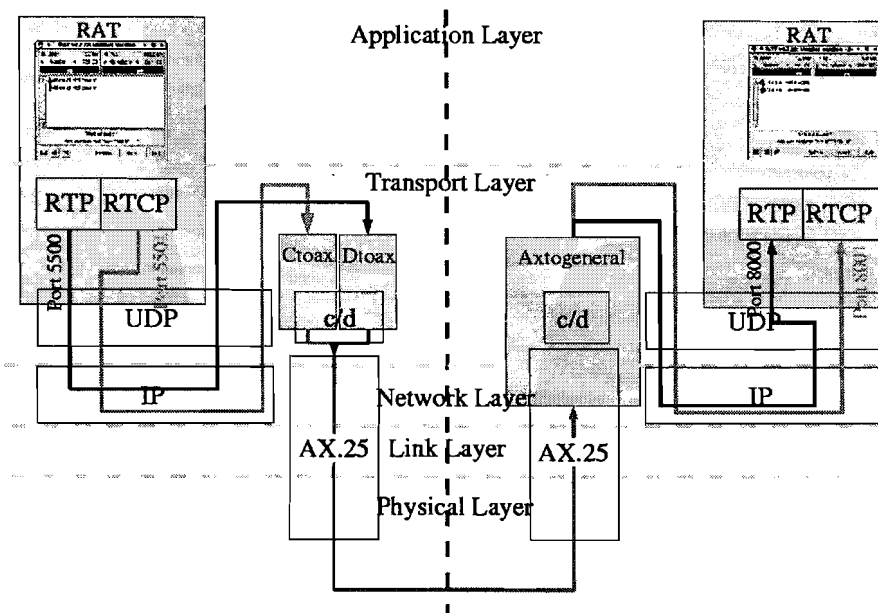
Figure 2: Overview of the speech transmission system

reports about the network quality parameters to the sender or to exchange extra information between the parties of a session eg. their email addresses or location.

Multiplexing the data streams of the speech and the control information requires extra functionality. In an IP network these streams are separated by different UDP port addresses, yet AX.25 does not support port multiplexing. Therefore an extra header of one octet had to be added to the packet which indicates whether there is control data or speech contained in the frame.

Thus the IP to AX.25 converter consists of 3 processes:

- one listening to the UDP port for RTCP control messages, generating AX.25 frames with control information

- one listening to the UDP port for RTP based speech transmission, generating AX.25 frames with speech data

- one listening to the AX.25 port, detecting whether control messages or speech data has arrived and resending the content to the appropriate UDP port.

5 Results

Several transmissions have been performed collecting speech samples for PESQ and subjective assessment.

Table 2 and Figure 3 present the results of the objective assessment of the collected samples. It can be seen that the range of loss percentage matches the results from the emulation satisfyingly. Therefore loss up to more than 50% and an average loss percentage of approximately 10% have to be accepted for real-time speech transmission using AX.25. This

Table 2: Objective assessment of AX.25 speech transmission

(a) Packet Loss Rates

min. loss	0.0%
max. loss	54.0%
av. loss	9.89%

(b) PESQ MOS

speaker	male	female
min. MOS	1.260	1.278
max. MOS	2.825	2.046
av. MOS	2.324	1.645

can hardly satisfy, as it will result in unrecognisable speech in many cases.

The speech quality calculation has been performed gender dependant, as the speech samples spoken by female speakers subjectively sounded poorer after having been transmitted. This effect of gender dependency has also been observed by [SI02] for ACELP codecs and is caused by the codec itself rather than the transmission.

The PESQ scores support the observations made from the packet loss rates: On the one hand there are cases in which the speech is considered to be intelligible and the upper bound of the speech quality achievable with the used codec is reached. On the other hand, MOS scores of almost 1 occur which indicate a complete distortion of the speech.

In average the male speech samples are at least recognisable even if the speech quality is not yet satisfying. For female speakers in average the recognisability of the speech cannot be guaranteed.

For subjective listening tests a subset of the collected speech samples has been chosen in such a way, that the most significant area of loss (0%-20%) was covered. For the chosen speech samples the PESQ MOS has been calculated as well to verify, that the samples cover approximately the same range of PESQ scores as above. These scores are presented in Table 3. The chosen speech samples have been
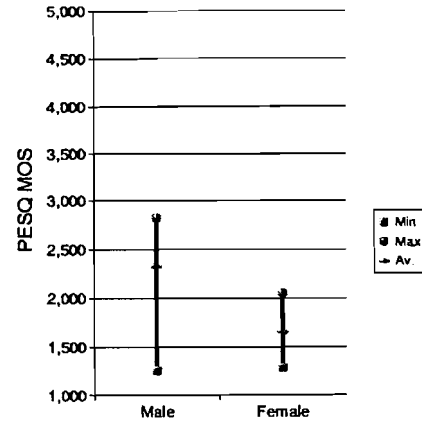

Figure 3: Objective assessment of AX.25 speech transmission

Table 3: PESQ MOS scores of the chosen subset

speaker	male	female
min. PESQ MOS	1.260	1.325
max. PESQ MOS	2.825	2.046
av. PESQ MOS	2.011	1.627

played to untrained subjects for listening tests using the P.800 [ITU96] MOS scale. The test conditions demanded by the ITU-Recommendation P.800 (eg. a sound-proof room) could not be fulfilled completely, however the results can give an idea about the expected speech quality.

Table 4: Subjective MOS scores for the chosen subset

speaker	male	female
min. MOS	2.167	1.778
max. MOS	3.389	3.056
av. MOS	2.880	2.321

Table 4 presents the results of this subjective test series: The subjects rated the speech samples much higher than expected from PESQ. The given marks where about one point above the PESQ scores, as can also be seen from Figure 4. Therefore the listeners showed a much higher tolerance to poor sounding voice than expected and rated the speech being at least comprehensible.
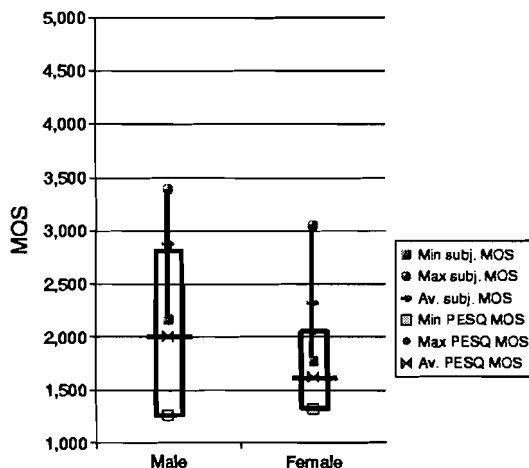

Figure 4: Comparison of PESQ with subjective MOS for the chosen subset

It is beyond the scope of this paper to discuss whether this difference is caused by inaccuracy of the objective or the subjective methods or both. For the aim of this study only a tendency for the expected speech quality had to be found proving the real-time abilities of AX.25 in general.

6 Conclusion

This paper has aimed to prove an AX.25 transmission system to be able to transmit digitised speech in real-time. It has found several drawbacks of packet radio for real-time transmission, especially a very large packet header which causes the need to produce comparable large packets to achieve data rates below 9.6 kBit/s. The codec used so far has shown a poor performance especially for female speakers causing a low speech quality even in loss-less cases which will occur seldom in wireless transmissions. Different tests with a transmission channel have found a very high loss percentage in some cases which cause very low speech quality. Still subjects listening to the resulting speech have shown a much higher tolerance to distortions than expected, so that the system fulfils the requirements of a first implementation.

However there are major improvements to do. In the first place a better codec has to be chosen to improve the overall speech quality independently of the loss percentage. In addition more modern codecs conceal loss much better than the chosen one, so that the speech quality can be expected to increase especially under lossy conditions. Furthermore experiments eg. with multiple-hop-topologies have to be taken into consideration as digipeating is one major point of AX.25. The system itself supports digipeating, still the impact of multiple hops on the speech quality has to be evaluated. Higher data rates than 9.6 kBit/s or a codec with higher compression (without reducing the speech quality further) have to be demanded for Voice over AX.25 as well: So far multiplexing of real-time data with other traffic is not possible as the whole bandwidth is needed for speech

transmission. Concurrent traffic will increase the packet loss and reduce the speech quality further.

To conclude, this study has shown, that AX.25 is able to transmit real-time speech under special conditions. Being only a first step, one can expect that improvements are possible leading to a system which delivers satisfying results.

References

[BNT97] William A. Beech, Douglas E. Nielsen, and Jack Taylor. *AX.25 Link Access Protocol for Amateur Packet Radio*, November 1997.

[CS00] Guo Chuanxiong and Zheg Shaoren. Analysis and Evaluation of the TCP/IP Protocol Stack of LINUX. In *Proceedings of the WCC 2000 congress*, August 2000.

[Eck99] Holger Eckardt. 70cm-FM/FSK-Transceiver T7F. web page: http://www.df2fq.de/produkte/T7F.html, June 1999. found sept 2002.

[ITU96] ITU. Methods of subjective determination of transmission quality. ITU-T Recommendation P.800, August 1996.

[Pal99] Nico Palermo. YAM Yet Another Modem. web page: http.//microlet.com/yam/, May 1999. found sept 2002.

[rat02] The Robust Audio Tool (RAT). webpage: http://www-mice.cs.ucl.ac.uk/multimedia/software/rat/, July 2002. found sept 2002.

[RBHH00] Antony W. Rix, John G. Beerends, Michael P. Hollier, and Andries P. Hekstra. PESQ - the new ITU standard for end-to-end speech quality assessment. In *Proceedings of the AES 109th convention, Los Angeles*, 2000.

[SCFJ96] H. Schulzrinne, S. Casner, R. Frederick, and V. Jacobson. RTP: A Transport Protocol for Real-Time Applications. RFC 1889, January 1996.

[SI02] L. Sun and E. C. Ifeachor. Perceived Speech Quality Prediction for Voice over IP-based Networks. In *Proceedings of the IEEE International Conference on Communications*, New York, April 2002. IEEE.

[SLHM01] H. Sanneck, N. T. L. Le, M. Haardt, and W. Mohr. Selective Packet Priorization for Wireless Voice over IP. In *Proceedings of the Fourth International Symposium on Wireless Personal Multimedia Communication*, Aalborg, Denmark, 2001.

[SWLI01] L. Sun, G. Wade, B. Lines, and E. C. Ifeachor. Impact of Packet Loss Location on Perceived Speech Quality. In *Proceedings of the Internet Telephony Workshop (IP-TEL)*, page 114 to 122, New York, 2001.