PERFORMANCE EVALUATION OF AMR WIDEBAND SPEECH TRANSMISSION FOR HANDS-FREE CAR KITS

Eric Böhmler, Jürgen Freudenberger, Michael Müller

Department of Computer Science
University of Applied Sciences Constance, Germany
phone: + (49) 7531 206 647, fax: + (49) 7531 206 559, email: juergen.freudenberger@htwg-konstanz.de
web: www.edc.in.htwg-konstanz.de

ABSTRACT

The adaptive multi-rate - wideband (AMR-WB) speech coding standard will improve the speech quality of telecommunication applications compared to the traditional narrowband speech transmission. This work presents a performance evaluation of the different modes of the AMR-WB standard for hands-free car kits. These investigations consider different noise situations, single and multichannel noise reductions systems, and the possible Bluetooth link between the hands-free device and the mobile phone.

1. INTRODUCTION

The AMR-WB speech codec is a speech coding standard which enables wideband audio transmission with a bandwidth in the range 50 Hz - 7 kHz [1]. This extension of the classical narrowband signal (300 Hz - 3.4 kHz) will significantly improve the speech quality transmission in future telecommunication applications. The publications [2, 3, 4, 5] provide a closer look at today's wideband speech and audio coding standards.

This paper considers the AMR-WB speech transmission for hands-free applications in a car environment. The AMR-WB codec has been developed for use in mobile radio environments where typical usage conditions may include both channel errors and high-level background noise. In the course of AMR-WB standardization, the codec was extensively tested. The experiments covered tests for clean speech and speech in four types of background noise including one scenario with car noise [6]. Under the considered conditions, the AMR-WB codec provides robust operation. For car noise Mean Opinion Scores (MOS) from 2.7 for the lowest bit rate codec (6.6 kps) up to a MOS value of 4.2 with a data rate of 12.65 kbps were reported [6].

However, the speech quality of a hands-free car kit is not only affected by car noise. Speech has to be picked up as directly as possible to reduce reverberation and to provide a sufficient signal-to-noise ratio, where the distance between microphone and speaker depends significantly on the position of the driver and therefore on the size of the driver. Furthermore, other noise sources like airflow from electric fans or car windows have to be considered. Good noise robustness requires the use of noise suppression techniques like spectral subtraction [7] for single microphone systems. Such noise reduction algorithms improve the signal-to-noise ratio, but they usually introduce undesired speech distortion.

Originally introduced as optional features connected by a wire to mobile phones, hands-free devices are now generally available with wireless technology. Typically they use Bluetooth as wireless link to the mobile phone. However, the speech transmission according to the current Bluetooth standard does not support the new wideband technology. Currently, the speech signal is transmitted as a 64 kbps PCM coded signal without additional speech coding. The bandwidth is therefore limited to the narrowband frequency range. The PCM signal provides a high quality input signal for the speech codec of the mobile network, because transcoding of lossy speech codecs is avoided. To overcome the problem of the missing wideband support for the Bluetooth link the ITU-T Focus Group on From/In/To Cars Communication II has suggested the introduction of the ITU-T G. 722 standard for Bluetooth speech transmission [8]. This codec works on a sample basis and therefore introduces low delay. On the other hand, transcoding from G.722 to AMR-WB has to be performed in the mobile phone. Such a transcoding of lossy speech codecs almost always introduces generation loss. Moreover, the Bluetooth transmission may introduce additional transmission errors that may further diminish the speech quality.

The introduction of wideband speech requires a doubling of the sampling rate compared with conventional narrowband speech transmission. For hands-free units the higher sampling rate significantly increases the computational complexity of the required speech processing like noise reduction and echo compensation. It is therefore interesting to evaluate the potential quality improvements of the new wideband speech transmission for realistic hand-free scenarios.

This work presents a performance evaluation of the AMR-WB codec for hands-free car kits, where we consider reverberation, different noise situations, single and multichannel noise reduction systems, and the possible Bluetooth link between the hands-free device and the mobile phone. The speech quality is evaluated by means of instrumental quality measures as well as informal listening tests.

2. SIMULATION MODEL

For the performance evaluation of wideband speech transmission we used the system model depicted in Fig. 1. The system model represents the simulation of hands-free car communication.

The research was supported by the Baden-Württemberg Ministry of Science, Research and the Arts.

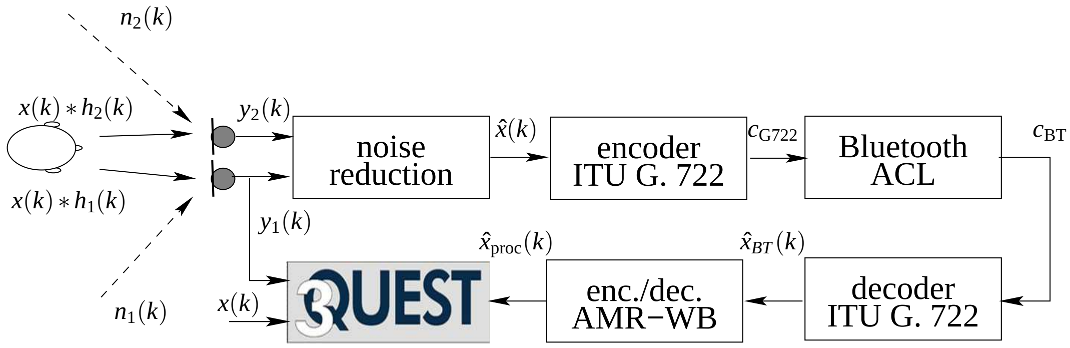

Fig. 1: System model

The clean speech signal is transmitted from the mouth of an artificial head to the hands-free microphones. Thereby the speech signal is distorted by car noise and reverberation. We assume that the acoustic system is linear, i.e. we can model the system by its impulse responses.

To improve the signal-to-noise ratio the unprocessed speech, recorded by the hands-free microphones, is processed by a noise reduction algorithm in the hands-free unit. The output signal of the hands-free unit is transmitted to the cellular phone over a Bluetooth link. To compress the speech signal for the Bluetooth transmission we assume that the ITU-T wideband speech codec G.722 is used as proposed in [8]. The encoded bits are transmitted using an asynchronous connectionless link (ACL).

The speech data is then transcoded by the AMR-WB codec which is utilized in the Universal Mobile Telecommunications System (UMTS). We assume that the transmission over the cellular network is error-free. The AMR-WB codec operates with 9 different bit rates from 6.6 to 23.85 kbps. For our investigations we considered only the bit rates 6.60, 12.65 and 23.85 kbps.

For the evaluation of the speech quality the decoded speech is processed by the 3QUEST algorithm. In the following subsections some system blocks are discussed in more detail.

2.1 Acoustic System

Our simulations are based on measurements with two cardioid microphones with positions suited for car integration: One microphone was installed close to the inside mirror. The second microphone was mounted at the A-pillar. With respect to three different background noise situations, we recorded driving noise at 100km/h and 140km/h. As third noise situation we considered the noise which arises from an electric fan (defroster). With an artificial head the impulse responses to the two microphones were measured.

As clean speech signals we took the German-speaking test samples from the recommendation P.501 of the International Telecommunication Union (ITU) [9]. For the simulation, the clean speech $x(k)$ was convoluted with the impulse responses $h_1(k)$ and $h_2(k)$ of the acoustic system. Then the recorded car noise $n_1(k)$ and $n_2(k)$ was added. The i^{th} microphone signal is therefore calculated as

$$y_i(k) = x(k) * h_i(k) + n_i(k)\,.$$

To improve the signal-to-noise ratio, hands-free units usually utilize noise suppression techniques. We consider to different acoustic front ends: a single microphone system and a two microphone noise reduction system. The single-channel algorithm uses the microphone installed close to the inside mirror and spectral subtraction to suppress ambient noise [10]. For the two channel approach the microphone signals are combined as proposed in [11]. The input and output signal-to-noise ratios for both algorithms are given in Table 1. Note that for the considered noise scenarios, the two channel noise suppression achieves an improvement of the MOS values of 0.5 compared with the single-channel algorithm (informal listening tests with respect to $\widehat{x}(k)$).

2.2 G.722 over Bluetooth

For the radio transmission between the hands-free unit and the cellular phone the Bluetooth ACL connection with a maximum bidirectional data rate of 432.6 kbps was used. The advantage of the ACL connection compared with the synchronous connection oriented (SCO) connection is that it allows to transmit a binary coded speech signal. Bluetooth supports circuit switched connection with a data rate of 64 kbps or 128 kbps. But the current specifications for SCO connections support only the transmission of PCM coded narrow band speech data with 64 kbps. The SCO link is therefore not suitable for a wideband transmission. The disadvantage of the ACL connection for speech transmission is that latency and delay increases compared to a SCO connection. Furthermore, the ACL connection is packet oriented and the timely arrival of speech packages is not guaranteed. To consider the effects of late package arrivals we used a statistic model as described in section 3.

To allow a wideband audio transmission over the ACL connection the speech signal was encoded with the ITU G.722 codec. The G.722 codec provides a signal bandwidth of 7000 Hz and reduces the bit rate to 48, 56, or 64 kbps, respectively. In our system model we delimit the bit rate to 64 kbps to achieve the best possible speech quality. After the simulated Bluetooth transmission the G.722 decoder reconstructs the encoded speech data.

Noise Reduction	Environment	SNR of $y_1(k)$	SNR of $y_2(k)$	SNR of $\widehat{x}(k)$
Two channel noise reduction	100 km/h	4.14	5.77	15.70
	140 km/h	-0.64	1.15	11.64
	Defroster	6.65	7.19	16.96
Single channel noise reduction	100 km/h	4.14	-	14.12
	140 km/h	-0.64	-	9.61
	Defroster	6.65	-	16.62

Tab. 1: Input and output signal-to-noise ratios for the applied noise reduction algorithms for three driving scenarios.

2.3 Evaluation of the Results

The speech quality of the wideband transmission has been evaluated by means of instrumental quality measures as well as informal listening tests. The instrumental quality analysis of the generated audio data was done by the tool 3QUEST [12]. It rates the signal by the mean opinion score (MOS) and therefore automates the subjective tests. 3QUEST takes as input files the clean speech signal $x(k)$ as a quality reference, the noisy signal $y_{1,2}(k)$ which arrives at the microphones and the processed speech $\hat{x}_{Proc}(k)$ which runs through the complete simulation chain. 3QUEST computes three MOS values to rate the speech quality.

N-MOS Quality of the noise signal.

S-MOS Quality of the speech signal.

G-MOS Overall quality of the speech signal.

Each MOS value is expressed in one number in the range from 1 to 5, 1 being the worst and 5 the best.

3. MODEL OF BLUETOOTH ACL TRANSMISSION

The encoded audio data c_{G722} is transmitted over a Bluetooth ACL connection between the handsfree unit and the cell phone. The errors which happen during this transmission can not be reproduced accurately for the same scenario in real Bluetooth transmissions. Therefore a simulation of the Bluetooth transmission was created. The simulation is driven by a model which reproduces errors which share the same characteristics as the errors which were recorded on the measurements of a scenario.

Two models were created from measurements in two different scenarios where the set up of the Bluetooth devices was arranged in different ways. The measurements of each scenario trained the models which are based on Markov chains. The model is able to estimate how many packets are going to be received from step to step depending on the state of the transmission.

Models were created for two scenarios where two Bluetooth devices were placed inside a car. The distance between both devices was approximately 1.5 m. With one scenario a Bluetooth device was additionally covered to shield the Bluetooth antenna.

Both machines were equipped with the same Bluetooth chip set. The data stream in each direction had a throughput of 64 kbps which is the flow-rate of the G.722 encoded audio stream. The stream was fragmented into packets of 80 byte which are sent in 10 ms time slots. Each receiver expects these packets in the same 10 ms but has a jitter buffer of 5 ms implemented. So the latency sums up to 15 ms. The measurements were done with a bidirectional transmission of ACL packets. The packet error rates for the two considered scenarios were 0.5% and 2%, respectively. ACL implements a retransmission protocol for a reliable connection and guarantees that all packets are received in the correct order. So in this application an error occurs if the packet was not received within the 15 ms. If no packets were received within a time slot a simple error concealment was done by repeating the last packet of encoded G.722 data. If multiple packets were received during one time slot the latest packet was taken. More advanced concealment methods are proposed in the appendices III and IV of the ITU-T G.722 recommendation.

An error consists typically of two phases. The first phase is when the connection is disturbed and the packet stream stalls. During this phase it is very likely that no packets are received over a time span which can last multiple time slots. After the connection is not disturbed any more the catch up phase begins. This phase can also last for several time slots. During each time slot of this phase one or typically more of the delayed packets are received.

To model the dependency of received packets during different phases the states of the Markov chain had to be split up into two characteristics. The first one gives the number of packets received in the last time slot. The second gives the number of packets the machine is behind of a synchronous transmission. The first characteristic describes the state of the transmission. The second characteristic determines the duration of the phase of an error. The longer the connection stalls the more likely it will recover and change to the catch up phase.

Both characteristics are combined in a table (see Fig. 2). Each entry represents a state. And each state points to a probability distribution which maps a probability to the number of packets which are expected to be received in the next time slot.

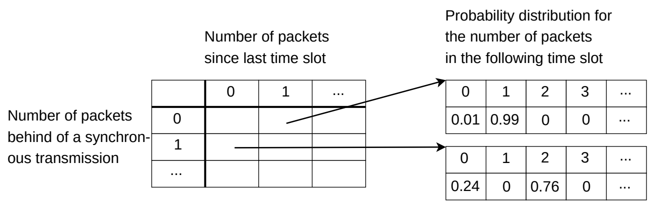

Fig. 2: Bluetooth model based on a Markov chain. The table to the left specifies the state of a transmission error. Each state points to a probability distribution of how many packets are expected in the next time slot.

Listening tests		AMR-NB	AMR-WB			AMR-WB & G.722
	AMR bitrate	12.20	6.60	12.65	23.85	23.85
Noise reduction	Environment	G-MOS	G-MOS	G-MOS	G-MOS	G-MOS
Two channel noise reduction	100 km/h	3.4	3.9	-	4.3	4.4
	140 km/h	2.4	3.5	-	3.5	3.6
Single channel noise reduction	100 km/h	2.1	-	-	3.0	-
	140 km/h	1.5	-	-	2.4	-

3QUEST results		AMR-NB	AMR-WB			AMR-WB & G.722
	AMR bitrate	12.20	6.60	12.65	23.85	23.85
Noise reduction	Environment	G-MOS	G-MOS	G-MOS	G-MOS	G-MOS
Two channel noise reduction	100 km/h	3.1	3.6	3.6	3.6	3.6
	140 km/h	2.3	3.0	3.0	3.1	3.1
Single channel noise reduction	100 km/h	3.1	3.4	3.4	3.4	3.5
	140 km/h	2.5	3.0	3.0	3.1	3.1

Tab. 2: MOS values from listening tests with 21 test persons (upper table) and the 3QUEST results (lower table), error-free Bluetooth transmission.

Noise reduction	Environment	AMR-WB 6.60			AMR-WB 23.85		
		S-MOS	N-MOS	**G-MOS**	S-MOS	N-MOS	**G-MOS**
Two channel noise reduction	100 km/h	4.2	3.6	3.6	4.3	3.4	3.6
	140 km/h	3.7	3.0	3.0	3.9	2.9	3.1
	Defroster	3.9	3.3	3.3	4.1	3.2	3.4
Single channel noise reduction	100 km/h	4.2	3.0	3.3	4.4	3.0	3.5
	140 km/h	4.1	2.4	3.0	4.3	2.4	3.1
	Defroster	4.1	3.1	3.3	4.3	3.0	3.4

Tab. 3: Detailed 3QUEST results for error-free Bluetooth transmission.

Noise reduction	Environment	AMR-WB 6.60			AMR-WB 23.85		
		S-MOS	N-MOS	**G-MOS**	S-MOS	N-MOS	**G-MOS**
Two channel noise reduction	100 km/h	3.9	3.6	3.4	4.0	3.4	3.4
	140 km/h	3.4	2.9	2.8	3.5	2.9	2.8
	Defroster	3.7	3.3	3.2	3.9	3.3	3.2
Single channel noise reduction	100 km/h	4.0	3.1	3.2	4.1	3.0	3.3
	140 km/h	3.8	2.7	2.9	3.9	2.7	3.0
	Defroster	4.0	3.2	3.3	4.2	3.0	3.4

Tab. 4: Detailed 3QUEST results for erroneous Bluetooth transmission.

4. SIMULATION RESULTS

Tab. 2 presents simulation results of the instrumental quality measures calculated by the 3QUEST algorithm as well as results of the listening tests with 21 test persons. The rightmost column contains the results with G.722 transcoding, where we assume error-free Bluetooth transmission. All other values were obtained without G.722 transcoding.

A notable result of the listening test is the subjective improvement of the voice quality from narrowband to wideband speech transmission. Tab. 2 shows that the test persons rated the wideband speech with the AMR-WB 23.85 codec about one point better than the narrowband signals in all considered background noise situations. Even with the data rate of 6.6 kbps, an improvement of 0.5 to 0.9 points is achieved. Comparing the results with and without G.722 transcoding, we observed almost the same voice quality in direct A-B comparisons. This is also indicated by the results in Tab. 2. Interestingly, the test persons rated the signals with lossy transcoding slightly better than the speech signals without transcoding. The presented results also demonstrate that the overall speech quality strongly depends on the employed noise reduction approach. The single channel scheme leads to more speech distortion and therefore lower MOS values.

Comparing the results of the listening tests with the values obtained with 3QUEST, we note that the 3QUEST values do not reflect the results of the listening tests in detail. The 3QUEST results, however, correctly predict trends in scenarios with the same noise reduction algorithm.

Tab. 3 and Tab. 4 display the MOS values as calculated by 3QUEST and compare an ideal Bluetooth transmission with the simulated erroneous transmission. The comparison shows a decline in the MOS values of about 0.2 points when transmission errors are considered. The two Bluetooth scenarios produced almost identical MOS values, so Tab. 4 represents both scenarios. Further tests in an office scenario showed a rapid decrease in the MOS values when the Bluetooth connection is disturbed, e.g. when the Bluetooth devices are separated over longer distances or by walls. The MOS values as calculated by 3QUEST dropped by one point when the devices were located 5 m from each other and separated by a wooden wall in comparison to a transmission where both devices were located next to each other.

5. CONCLUSIONS

In this work we have presented a performance evaluation of the AMR-WB standard for hands-free car kits.

The results of the instrumental quality measures as well as the listening tests show that the AMR-WB codec increases the speech quality significantly in comparison to the AMR-NB codec. This is true for all considered car noise scenarios. The AMR-WB codec displays good noise robustness in combination with different noise suppression techniques.

For in-car applications with erroneous Bluetooth transmission, the simulation results show that the G.722 in combination with the Bluetooth ACL connection is robust against transmission errors. Transcoding to the G.722 codec has little influence on the speech quality as long as the Bluetooth connection introduces few transmission errors. Therefore, the G.722 codec is a suitable candidate for introducing wideband speech transmission in the Bluetooth standards. However, in situations where the Bluetooth devices are separated over longer distances or by walls, a rapid decrease of the MOS values was experienced. Here, an improvement could possibly be achieved by disabling the retransmission algorithm of the ACL connection and using a best effort transmission instead. This is the subject of further investigations.

REFERENCES

[1] 3GPP TS 26.190, *Adaptive Multi-Rate wideband speech transcoding*, 3GPP Technical Specification.

[2] R.V. Cox, S.F. De Campos Neto, C. Lamblin, and M.H. Sherif, "ITU-T coders for wideband, superwideband, and fullband speech communication," *Communications Magazine, IEEE*, vol. 47, no. 10, pp. 106 –109, october 2009.

[3] P. Jax P. Vary, "Bandwidth extension of speech signals: A catalyst for the introduction of wideband speech coding?," *IEEE Communications Magazine*, vol. 1, pp. 106–111, 2006.

[4] P. Vary H. Taddei S. Schandl M. Gartner C. Guillaume S. Ragot B. Geiser, P. Jax, "Bandwidth extension for hierarchical speech and audio coding in ITU-T Rec. G.729.1," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 15, pp. 2496–2509, 2007.

[5] Ronald M. Aarts Erik Larsen, *Audio Bandwidth Extension*, Wiley, 2004, Application of Psychoacoustics, Signal Processing and Loudspeaker Design.

[6] 3GPP TR 26.976, *AMR-WB Speech Codec Performance Characterization*, 3GPP Technical Specification.

[7] S. Boll, "Suppression of acoustic noise in speech using spectral subtraction," *IEEE Trans. Acoustics, Speech, and Signal Processing*, vol. 27, pp. 113–120, 1979.

[8] Focus Group on From/In/To Cars Communication II, *Requirements on Wideband Bluetooth Connection from FG CarCom Point of View*, ITU-T, 2008.

[9] ITU-T, "Test signals for use in telephonometry," Recommendation ITU-T P.501, International Telecommunication Union, Geneva, 2007.

[10] R. Martin, "Spectral subtraction based on minimum statistics," in *Proc. Eur. Signal Processing Conf.*, 1994, pp. 1182–1185.

[11] J. Freudenberger, S. Stenzel, and B. Venditti, "Spectral combining for microphone diversity systems," in *Proc. Eur. Signal Processing Conf. (EUSIPCO-2009), Glasgow*, 2009.

[12] H.W. Gierlich, F. Kettler, S. Poschen, and J. Reimes, "A new objective model for wide- and narrowband speech quality prediction in communications including background noise," in *16th European Signal Processing Conference (EUSIPCO 2008)*, 2008.