Proceedings of 20th International Congress on Acoustics, ICA 2010

23–27 August 2010, Sydney, Australia

Do We Need Dereverberation for Hand-Held Telephony?

Marco Jeub[1], Magnus Schäfer[1], Hauke Krüger[1], Christoph Nelke[1], Christophe Beaugeant[2], and Peter Vary[1]

[1] Institute of Communication Systems and Data Processing (ind), RWTH Aachen University, Germany
[2] Infineon Technologies, Sophia-Antipolis, France
{jeub,schaefer,krueger,nelke,vary}@ind.rwth-aachen.de, christophe.beaugeant@infineon.com
www.ind.rwth-aachen.de   www.infineon.com

PACS: 43.55.Hy, 43.72.Kb, 43.72.Gy

ABSTRACT

The necessity of dereverberation algorithms in hand-held speech communication systems is discussed in this contribution. The study is based on a new measurement campaign with artificial head and a two-microphone mock-up phone in realistic acoustical environments like an office, corridor and stairway. Based on objective speech quality measures as well as a listening test, we show that room reverberation can lead to a decrease in intelligibility even for hand-held telephony under certain conditions. Hence, the far-end listener can benefit from dereverberation algorithms in the sending device. All measured room impulse responses are available online as part of the Aachen Impulse Response (AIR) database.

1. INTRODUCTION

In speech communication systems, room reverberation often leads to a degradation of speech quality and intelligibility. This applies for hands-free devices and digital hearing aids. In contrast to that, the influence of reverberation on the intelligibility in hand-held telephony is commonly assumed to be negligible. In this contribution we show that this statement is not always true.

The common tool for the evaluation of speech codecs in reverberant environments is the ITU-T G.191 Software Tool Library (STL) [1]. It includes a reverberation module consisting of different room impulse responses (RIR). These were measured in an office and meeting room with a direct loudspeaker-microphone path. This however, is an unrealistic assumption for phone conversations in the usual Hand-Held Position (HHP) [2]. The resulting indirect sound propagation from the mouth to the microphones at the device as well as diffraction effects of the head are not taken into account. Additionally, no standardized positions were used to generate the STL impulse responses.

For the evaluation of dereverberation algorithms for (binaural) digital hearing aids, we recently published the Aachen Impulse Response (AIR) database [3], covering a wide range of realistic situations. So far, this database contains no measurements suitable for the evaluation of speech transmission systems.

The aim of this contribution is twofold. First, we present new measured RIRs which can be seen as an extension to both STL and AIR. Second, we give an elaborate study of reverberation effects in hand-held telephony based on objective measures and a listening test using the measured RIRs in combination with different speech codecs (narrowband, wideband, superwideband). All measurements are available online as an extension of the AIR database [1].

The remainder is organized as follows. In Section 2 we investigate the ITU-T Software Tool Library and clarify the drawbacks of the included RIRs. Section 3 describes the measurement procedure of the new RIRs including measurement rooms and recording soft- and hardware. Section 4 explains our system for simulating speech transmission. Section 5 analyses the measured impulse responses in terms of acoustical properties followed by subjective evaluations including the results of a listening test in Section 6. Finally, in Section 7 we draw some conclusions.

[1]The Aachen Impulse Response (AIR) database can be found at http://www.ind.rwth-aachen.de/AIR

2. ITU-T G.191 REVERBERATION MODULE

The ITU-T G.191 Software Tool Library (STL) is a free library for speech and audio coding standardization [1]. It consists of several speech codecs (e.g. G.711, G,726, G.722), measurement tools (e.g. speech activity meter) and transmission impairment functions (e.g. error insertion device). In 2005, the STL was, among other features, extended by a reverberation module. Besides a convolution tool, three different room impulse responses are provided. They were measured in a small video-conferencing room and an office room. Table 1 shows the corresponding loudspeaker-microphone distances d_{LM}, reverberation times (RT) and direct-to-reverberant energy ratios (DRR). Definitions of RT and DRR will be explained later.

Room	d_{LM} (m)	RT (s)	DRR (dB)
Office	0.5	0.34	18.64
Office	1.0	0.25	17.40
Conference room	0.5	0.20	16.55

Table 1: Acoustic properties of the STL room impulse responses.

Even though the STL reverberation module was designed for the development of speech and audio codecs, the provided RIRs show several drawbacks when it comes to the evalua-

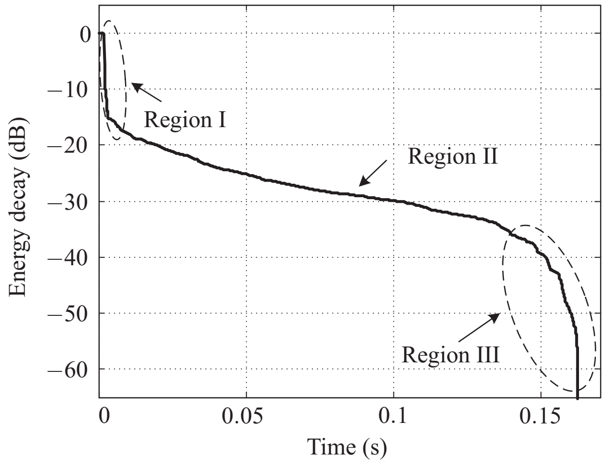


Figure 1: Energy decay curve of STL RIR, measured in an office room at $d_{\mathrm{LM}} = 0.5\,\mathrm{m}$.

tion of such algorithms under realistic conditions. All RIRs were measured without artificial head and a direct loudspeaker-microphone path. This however, is an unrealistic assumption for phone conversations in the usual hand-held position. The resulting indirect sound propagation from the mouth to the (possibly multiple) microphones at the device as well as diffraction effects of the head are not taken into account. Additionally, no standardized positions were used to generate the impulse responses.

Regarding the energy decay curve (EDC), exemplarily for the office room ($d_{\mathrm{LM}} = 0.5\,\mathrm{m}$) in Figure 1, the curve can be analyzed with respect to three different regions:

- Region I: Energy decrease between direct and reverberant components
- Region II: Sound decay
- Region III: Abrupt decrease of sound decay.

While Regions I and II are typical for such impulse responses, Region III indicates a cropping of the impulse response in its decay phase. This results in a high direct energy compared to the reverberant energy (higher DRR). It will be shown later that such high DRR values are not realistic for both HHP and Hands-Free Reference Point (HFRP) case. The corresponding office room impulse response ($d_{\mathrm{LM}} = 0.5\,\mathrm{m}$) of the STL is scaled to emphasize the different regions in Figure 2.

Hence, we conclude that the STL is not appropriate for a realistic evaluation of speech processing algorithms in realistic reverberant environments because of the direct loudspeaker-microphone path, high DRR as well as the abrupt decrease in the energy decay. This leads to an unnatural sound compared to a RIR with the full decay phase. Additionally, the STL neglects head shadowing and contains only impulse responses with a moderate reverberation time.

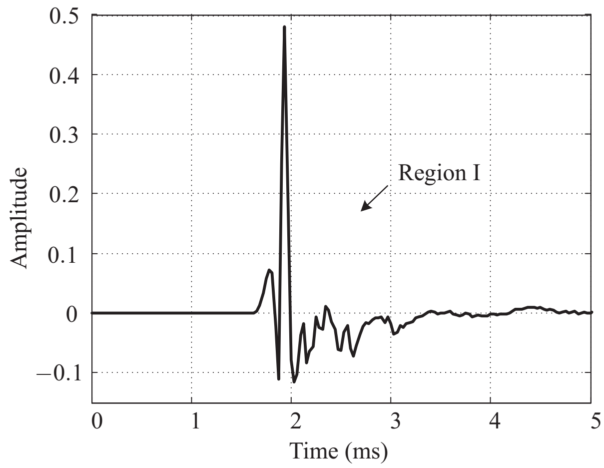
3. MEASUREMENT PROCEDURE

3.1 Measurement System

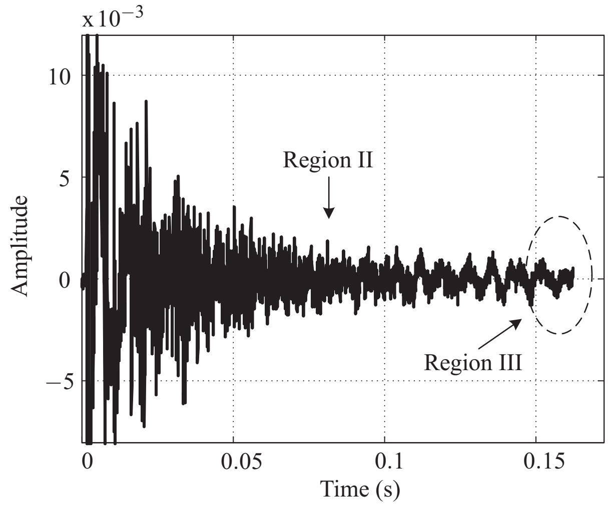
In order to capture impulse responses which can be used to simulate a realistic phone conversation, we measured such impulse responses in different rooms and positions of the phone. The recording systems consists of a HEAD acoustics HMS II.3 artificial head according to ITU-T Rec. P.58 [4] including a mouth simulator.

A mock-up phone was build out of two omnidirectional Beyerdynamic MM1 measurement microphones integrated in a

$6\mathrm{x}12\mathrm{x}3\,\mathrm{cm}^3$ plastic housing. A second microphone is used in order allow for an evaluation of dual-channel algorithms as well. The two microphones are placed with a 2 cm spacing in front of the mock-up phone. For measuring impulse responses



(a) Region I of the STL RIR.



(b) Regions II and III of the STL RIR.

Figure 2: Impulse response of the STL office room, scaled to emphasize (a) Region I and (b) Regions II and III.

in the HHP, the phone was mounted on the artificial head by means of the HEAD acoustics HHP 3 hand-held positioner in the flat handset position in accordance with ITU-T P.64 Annex D.3 [2], see Figure 3. Additionally, RIRs are measured with the phone placed at the Hands-Free Reference Point (HFRP) [5]. The measurements were performed with a standard laptop equipped with the RME Multiface II audio interface in combination with the RME Octamic II microphone amplifier. All measurements were performed with a sampling frequency of 48 kHz and 24-bit accuracy. For all measurement setups the sound pressure level (SPL) for the excitation signal was set to 89 dB at the mouth reference point (MRP).

3.2 Measurement Rooms

In order to investigate the influence of room reverberation in typical cell-phone conversations, we measured impulse responses in several realistic indoor environments:

- Office
- Kitchen
- Corridor
- Stairway
- Lecture room
- Meeting room

For the sake of brevity, we restrict our analysis to office, kitchen, corridor and stairway in the following. However, all measured RIRs are included in the AIR database. Dimensions of the dif-

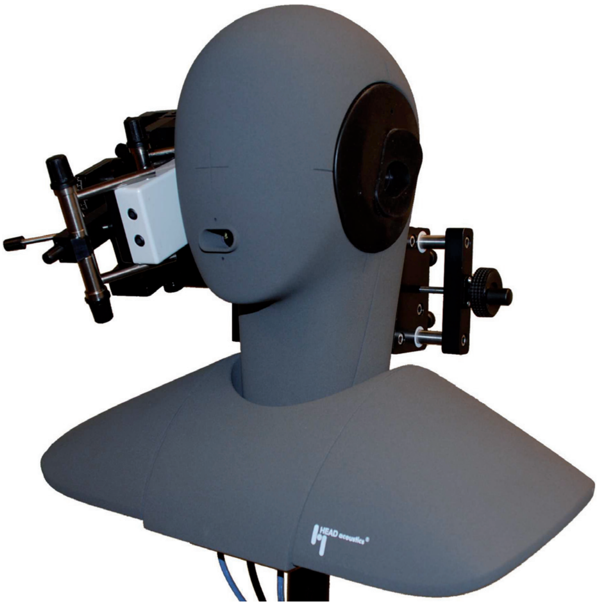


Figure 3: HEAD acoustics HMS II.3 artificial head with the two-microphone mock-up phone clamped in the HEAD acoustics HHP 3 hand-held positioner.

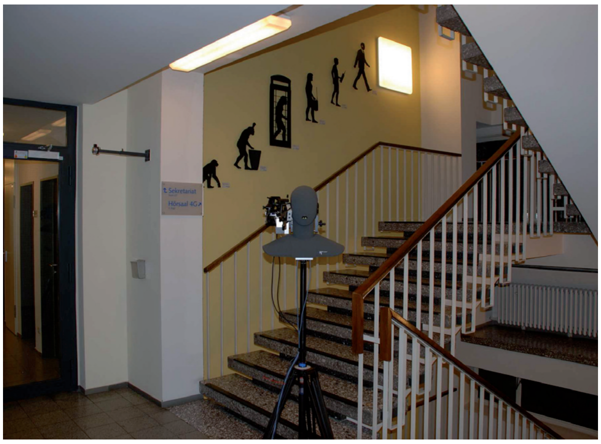


Figure 4: Measurement setup for the stairway.

ferent rooms can be found in Appendix A of this paper. Exemplarily, the picture of the stairway measurement setup can be found in Figure 4. All further rooms are shown on the corresponding website.

3.3 Measurement Software

The software to measure the impulse responses runs on a laptop using Windows 7 and Matlab r2009b. The time critical synchronous playback and recording of audio data is realized by RTProc [6] which is operated in the background of Matlab.

As a measurement signal, we used the recently introduced perfect sweeps (PSWEEP) [7]. Perfect sweeps are a new class of real-valued perfect sequences being similar to traditional sine sweeps but having a perfectly constant magnitude spectrum. They incorporate the benefits of so-called perfect sequences and traditional linear sine sweeps. Additionally, when used repeated periodically, the perfect sweep and all its derivatives are continuously differentiable especially at the edges between two successive repetitions. This prevents clicking noise and makes this signal particularly suitable for the use with acoustical measurement hardware.

4. SPEECH TRANSMISSION FRAMEWORK

The aim of this contribution is to evaluate the effects of room reverberation on speech communication systems. Therefore, we employ several speech codecs to the reverberant speech signal in order to generate realistic transmitted speech signals. These signals are for performed the listening test.

In the context of CELP speech codecs, it is already known that the effects of room reverberation are reduced by means of the adaptive postfilter [8] employed in the speech decoder [9]. However, a sufficient dereverberation cannot be obtained by such processing, especially since the postfilter is employed at the receiver side. Additionally, most codecs do not employ the long-term postfilter proposed in [8], which performs the highest amount of reverberation reduction, for complexity reasons.

The test signals for our experiments are generated as follows. First, speech files $s(k)$ from the TSP speech database [10] are convolved with the impulse responses $h(k)$ between artificial mouth and microphones of the mock-up phone at $f_s = 48\,\text{kHz}$, yielding the reverberant signals $x(k)$. Second, the reverberant speech signals $x(k)$ are downsampled, encoded and decoded independently using three different speech codecs with sampling frequency, bandwidth and bit rates as follows:

- Adaptive multi-rate narrowband codec (AMR-NB) [11] $f_s = 8\,\text{kHz}$, 3.4 kHz, 12.2 kbit/s
- Adaptive multi-rate wideband codec (AMR-WB) [12] $f_s = 16\,\text{kHz}$, 7 kHz, 23.05 kbit/s
- Super-wideband (SWB) speech and audio codec [13] $f_s = 32\,\text{kHz}$, 14 kHz, 64 kbit/s

The reverberant and codec signals are denoted by $\tilde{x}(k)$ in the following. For simplicity, no bit errors were added.

5. OBJECTIVE EVALUATION

5.1 Impulse Response Analysis

In a first evaluation step the measured impulse responses are evaluated by comparing energies contained in different parts of the impulse responses as well as frequency independent and subband reverberation times. Such channel-based measures are given in the next subsection, followed by an analysis of the coherence between the two microphones of the mock-up phone.

5.1.1 Channel-Based Measures

All channel-based measures

- RT: Reverberation time
- DRR: Direct to reverberant energy ratio
- ETR: Early to total sound energy ratio
- ELR: Early to late reverberation ratio (Clarity index)

refer to the acoustic channel and are calculated directly from the given impulse responses. Table 2 shows the results for the measured rooms. The definition and an elaborate discussion of the energy-based channel measures can be found, e.g., in [14]. The reverberation time is defined as the time period a switched off sound need to decrease by 60 dB and is measured with the Schroeder method [15] by a least square fitting of the EDC between -5 and -30 dB.

It can be seen in Table 2 that the objective measures differ greatly between HHP and HFRP. This can be explained with the direct path between loudspeaker and microphone at the HFRP and the indirect sound propagation for the HHP. For both office and kitchen, a moderate reverberation time of < 0.5s was measured and no significant difference in the RT between HHP and HFRP was examined. However, the DRR differs by more than 6 dB between HHP and HFRP for all measured rooms. Regarding the corridor and stairway scenario, even for the HHP a high RT was measured.

When comparing the results of Table 2 with the corresponding measures for the STL in Table 1, it can be seen that the values exhibit significant changes. Especially the DRR shows

Room	RT (s)		DRR (dB)		ETR (%)		ELR (dB)	
	HHP	*HFRP*	*HHP*	*HFRP*	*HHP*	*HFRP*	*HHP*	*HFRP*
Office	0.39	0.52	11.79	5.12	99.17	94.83	24.45	15.90
Kitchen	0.42	0.52	11.18	4.62	98.92	93.53	22.92	15.06
Corridor	1.25	1.47	10.98	4.35	97.56	89.06	17.49	10.45
Stairway	0.86	1.23	12.28	6.41	98.91	93.68	21.69	13.90

Table 2: Channel-based measures calculated directly from the impulse responses. The results are averaged over both channels.

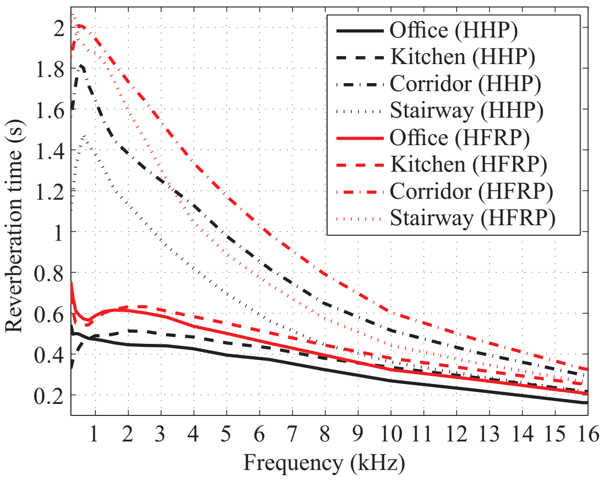


Figure 5: Subband reverberation time (SRT) calculated by means of the Schroeder method in $1/3$-octave subbands.

higher values for the STL impulse responses compared to the measured RIRs for both HHP and HFRP. Additionally, the measured RTs for the STL are quite moderate and no significant influence in the intelligibility results, especially since a very high amount of direct speech arrives at the microphones.

Since most rooms do not have a constant reverberation time over frequency, a reverberation time in subbands (SRT) is another significant parameter. Figure 5 shows the reverberation time as a function of frequency calculated by means of the Schroeder method and a $1/3$-octave filterbank. All measured impulse responses show an increasing RT for decreasing frequencies. While the curves are quite flat for office and kitchen, a more distinctive slope can be observed for the corridor and stairway scenario.

5.1.2 Coherence Analysis

The sound field in a reverberant room can be approximated by a diffuse sound field, cf. [14]. This has been shown in experiments in different acoustical environments with a loudspeaker-microphone distance of $d_{\mathrm{LM}} > 1\,\mathrm{m}$ in [3]. In this section it will be investigated if this approximation is also valid if the desired signal is captured by a telephone in HHP and HFRP position. In the following $x_1(k)$ and $x_2(k)$ represent the two microphone signals of the mock-up phone.

The magnitude squared coherence (MSC) between the two signals $x_{1|2}(k)$ is defined as

$$\Gamma^2_{x_1x_2}(\Omega) = \frac{\Phi^2_{x_1x_2}(e^{j\Omega})}{\Phi_{x_1x_1}(e^{j\Omega}) \cdot \Phi_{x_2x_2}(e^{j\Omega})}, \qquad (1)$$

where $\Phi_{x_1x_1}(e^{j\Omega})$ and $\Phi_{x_2x_2}(e^{j\Omega})$ represent the auto-power spectral densities (APSD) of $x_1(k)$ and $x_2(k)$, respectively. The cross-power spectral density (CPSD) between $x_1(k)$ and $x_2(k)$ is denoted by $\Phi_{x_1x_2}(e^{j\Omega})$.

The MSC between two microphones of an ideal spherically isotropic (diffuse) sound field can be expressed as [14]

$$\Gamma^2_{x_1x_2,\mathrm{diff}}(f) = \mathrm{sinc}^2\left(\frac{2\pi f d_{\mathrm{mic}}}{c}\right), \qquad (2)$$

with distance d_{mic} between two omnidirectional microphones and f denoting the frequency.

In the following, a division of the speech signal into its direct and reverberant components is performed. For the sake of brevity, the decomposition is given for one channel only. The decomposed input signal $x(k)$ can be expressed by

$$x(k) = \underbrace{\sum_{n=0}^{\mathrm{T}_d f_s - 1} s(k-n)h(n)}_{x_{\mathrm{d}}(k)} + \underbrace{\sum_{n=\mathrm{T}_d f_s}^{\mathrm{T}_r f_s} s(k-n)h(n)}_{x_{\mathrm{r}}(k)}, \qquad (3)$$

where the time span of the direct sound (including sound propagation) is given by T_d. Since we assume a high portion of direct sound, we can simply determine T_d by the global maximum of the RIR plus a few reflections (here: $2\,\mathrm{ms}$).

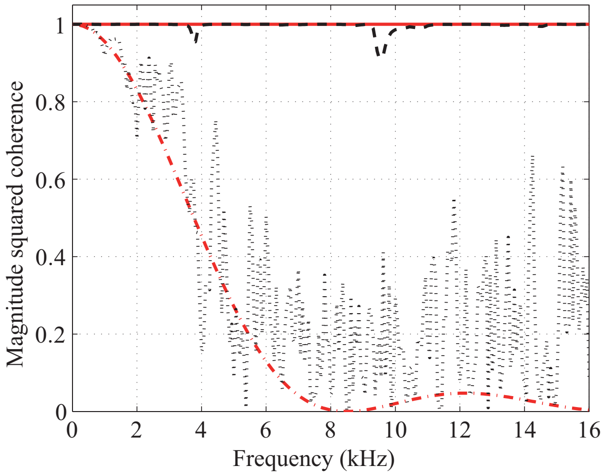
The MSC for direct and reverberant speech will be calculated as follows. First, two measured room impulse responses are decomposed into direct and reverberant components. Afterwards, speech data of $18\,\mathrm{s}$ duration from the TSP database is convolved with each of the RIRs resulting in separate direct and reverberant signals for each channel. Finally, the MSC between the two channels is calculated for both direct and reverberant speech by the Welch periodogram approach [16] using the Matlab command `mscohere`.

The corresponding curves for the stairway and office room are depicted in Fig. 6 (a) and (b). The upper dashed line shows the MSC of the direct speech component while the lower dashed curve shows the MSC of the reverberant speech component. The solid lines give the corresponding theoretical coherence function. As a high amount of direct speech is assumed, the theoretical coherence for the direct speech is one for all frequencies. The lower solid line gives the theoretical curve for an ideal diffuse sound field according to Eq.(2). It can be seen that the reverberant sound field for both the HHP and the HFRP scenario, is a good approximation of a diffuse sound field. Experiments with the other measured RIRs led to the same results. It has to be mentioned that, when using coherence-based algorithms which require a low coherence for the interference, noise attenuation can only be obtained for higher frequencies due to the small microphone spacing.

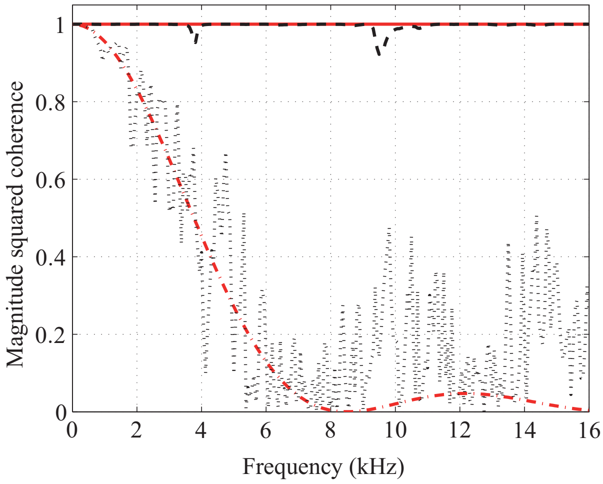
From the channel-based measures we can already conclude that room reverberation has an influence on the speech quality and intelligibility. Especially for stairways and corridors, the RT is very high even for the HHP. In the following subsection, a signal-based measure is evaluated which shows the degradation in speech quality.

5.2 Signal-Based Measures

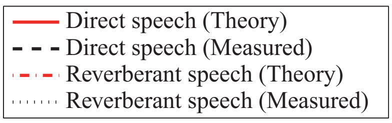
To rate the amount of speech distortion introduced by room reverberation, several objective measures exist in the literature. However, most of these measures do not show a high correlation with the subjective rating of a corresponding listening



(a) Office Room (HHP).



(b) Stairway (HHP).



(c) Legend.

Figure 6: Magnitude squared coherence (MSC) of direct and reverberant speech components for (a) the office in HHP position and (b) the stairway in HHP position.

Figure 7: Objective evaluation of the reverberant speech signal using the non-intrusive PEMO-Q measure. A PEMO-Q score of 1 indicates the highest perceptual similarity.

test. Investigations on such measures can be found, e.g., in [17]. Here, we restrict our analysis to the PEMO-Q measure (PSMt) [18] which was shown to be suitable for rating the subjective quality of reverberant and dereverberated speech [17]. The measure is based on an auditory model and calculates a perceptual similarity measure between the anechoic and degraded speech signal.

The results in terms of PEMO-Q (range: 0 to 1) are depicted in Figure 7. The figure shows the perceptual similarity measures between anechoic speech $s(k)$ and reverberant speech $x(k)$ and confirms the results of the channel-based evaluations. For rooms with moderate reverberation (Kitchen/office), a degradation can be observed for the hand-held position and even stronger for the hands-free case. In terms of stairway and corridor situations, room reverberation has a very strong influence for both HHP and HFRP.

The assumptions made by objective measures will now be confirmed by a listening experiment described in the next section.

6. SUBJECTIVE EVALUATION

6.1 Listening Test

The listening test took place in a low-reverberant studio booth having a high sound isolation of 42 dB against exterior noise. In order to ensure high quality audio and to avoid distortions due to the headphone, a calibrated HEAD Acoustics PEQ V digital equalizer in combination with a Sennheiser HD600 headphone was used.

During the test with 30 experienced listeners (normal hearing, age: $24-33$ years), 24 different signals $\tilde{s}(k)$ were presented to the participants. An anechoic speech signal of 18 s duration was processed according to the transmission system described in Section 4. Each of the 8 reverberant signals (4x HHP, 4x HFRP) was coded with the NB, WB and SWB codec.

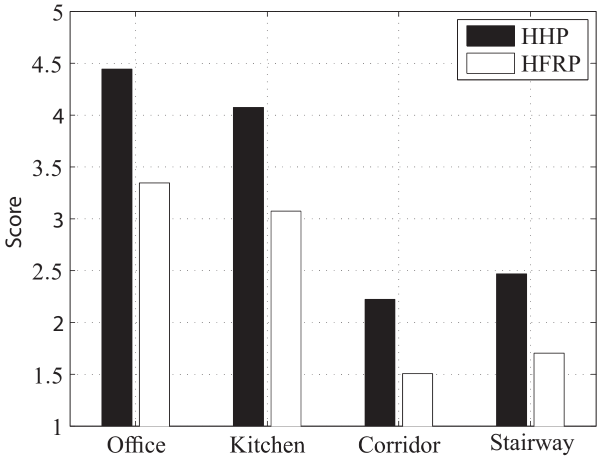
For each of the sentences, the listeners were asked to rate the impairment according to the ITU-R BS.1284-1 five-grade impairment scale (see Table 3 and [19]). The signals could be played ad libitum before the probands had to make their judgments. It has to be mentioned that the listeners were not asked to rate the overall speech quality but only the impairment due to room reverberation. Therefore, the results of the different codecs do not represent a quality rating.

5.0	Imperceptible
4.0	Perceptible but not annoying
3.0	Slightly annoying
2.0	Annoying
1.0	Very annoying

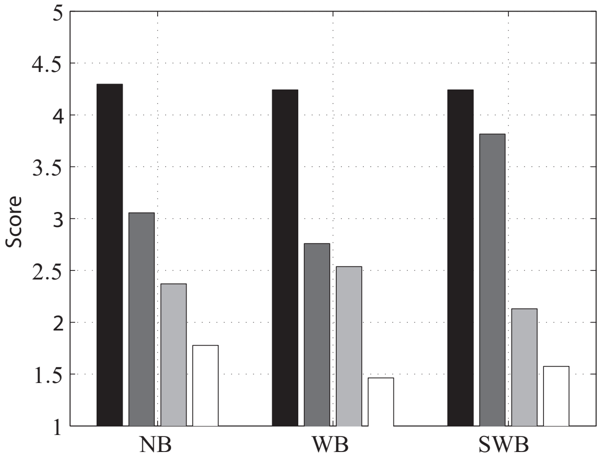
Table 3: ITU-R BS.1284-1 five-grade impairment scale [19].

The results averaged over the scores of the 30 participants and over the three codecs are depicted in Figure 8 (a). As can be seen from the figure, reverberation is perceptible for all tested scenarios. However, it can also be observed that there are some differences to the PEMO-Q results. While PEMO-Q indicates that reverberation has a strong influence for all rooms, the listening test shows that most listeners rated the effect for office and kitchen as perceptible but not annoying in the hand-held case. In terms of the corridor and stairway sentences, the effects of reverberation are clearly perceptible and rated as annoying. As expected, the impairment scores for the hands-free positions are always lower.

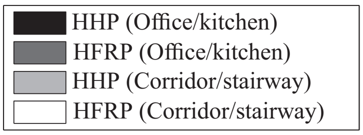
In a second experiment the subjective scores for the different



(a) Subjective score per room and position, averaged over all codecs.



(b) Subjective score for the different codecs, averaged over office/kitchen and corridor/stairway.



(c) Legend of (b)

Figure 8: Results of the listening test according to the ITU-R five-grade impairment scale: (a) averaged over the different codecs and (b) averaged over two sets of rooms (office/kitchen and corridor/stairway).

codecs are evaluated. The results are separated among rooms with moderate and strong reverberation and are shown in Figure 8 (b). It can be observed that no significant difference exists among the tested codecs, even though the SWB codec shows a lower reverberation influence for the office/kitchen HFRP case. This corresponds to the investigations in [20], where different wideband codecs where investigated under reverberant conditions.

7. CONCLUSIONS

In this contribution it has been analyzed if room reverberation has an impact on speech communication systems with hand-held telephony and hence, if dereverberation is needed.

The results of the measurements with artificial head and mock-up phone in realistic scenarios show that the objective channel-based measures RT and DRR exhibit significant differences compared to the ITU STL impulse responses, i.e., higher RT and lower DRR. Hence, the impulse responses provided in this contribution allow for a more realistic evaluation of a speech transmission system.

Based on the objective PEMO-Q measure as well as a listening test with 30 participants, it has been shown that an impairment due to reverberation can always be observed. For small enclosures like the tested office and kitchen, the effects are mostly not rated as annoying. But since room reverberation has a strong influence on the intelligibility in stairways and corridors, we conclude that dereverberation algorithms should be applied for both hands-free and hand-held telephones.

ACKNOWLEDGMENT

The authors would like to thank Aulis Telle for providing the perfect sweep generation software. Additionally, we like to thank Bernd Geiser for valuable contributions to the AIR database.

REFERENCES

[1] ITU-T Rec. G.191, *Software tools for speech and audio coding standardization*, ITU, 2005.

[2] ITU-T Rec. P.64, *Determination of sensitivity/frequency characteristics of local telephone systems*, ITU, 2007.

[3] M. Jeub, M. Schäfer, and P. Vary, "A binaural room impulse response database for the evaluation of dereverberation algorithms," in *Proc. Int. Conference on Digital Signal Processing (DSP)*, Santorini, Greece, 2009.

[4] ITU-T Rec. P.58, *Head and Torso Simulator for Telephonometry*, ITU, 1996.

[5] ITU-T Rec. P.340, *Transmission characteristics and speech quality parameters of hands-free terminals*, ITU, 2000.

[6] H. Krüger and P. Vary, "RTPROC: A system for rapid real-time prototyping in audio signal processing," in *Proc. 12th IEEE/ACM International Symposium on Distributed Simulation and Real-Time Applications DS-RT*, Vancouver, Canada, 2008, pp. 311–314.

[7] A. Telle, C. Antweiler, and P. Vary, "Der perfekte Sweep - Ein neues Anregungssignal zur adaptiven Systemidentifikation zeitvarianter akustischer Systeme," in *Proc. of German Annual Conference on Acoustics DAGA*, Berlin, Germany, 2010.

[8] J.-H. Chen and A. Gersho, "Adaptive postfiltering for quality enhancement of coded speech," *Speech and Audio Processing, IEEE Transactions on*, vol. 3, no. 1, pp. 59–71, 1995.

[9] M. Jeub and P. Vary, "Enhancement of reverberant speech using the CELP postfilter," in *Proc. IEEE Int. Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Taipei, Taiwan, 2009, pp. 3993–3996.

[10] P. Kabal, "TSP speech database," Tech. Rep., Department of Electrical & Computer Engineering, McGill University, Montreal, Quebec, Canada, 2002.

[11] 3GPP TS 26.071, *Adaptive Multi-Rate (AMR) speech codec; General description*, 3GPP, 2004.

[12] 3GPP TS 26.171, *Adaptive Multi-Rate - Wideband (AMR-WB) speech codec; General description*, 3GPP, 2004.

[13] B. Geiser, H. Krüger, H.W. Löllmann, P. Vary, D. Zhang, H. Wan, H. T. Li, and L. B. Zhang, "Candidate proposal for ITU-T super-wideband speech and audio coding," in *Proc. IEEE Int. Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Taipei, Taiwan, 2009, pp. 4121–4124.

[14] H. Kuttruff, *Room Acoustics*, Spon Press, Oxon, 2009.

[15] M.R. Schroeder, "New method of measuring reverberation time," *J. Acoust. Soc. Am.*, vol. 37, no. 3, pp. 409–412, 1965.

[16] P. Welch, "The use of fast fourier transform for the estimation of power spectra: A method based on time averaging over short, modified periodograms," vol. 15, no. 2, pp. 70–73, 1967.

[17] S. Goetze, E. Albertin, M. Kallinger, A. Mertins, and K.-D. Kammeyer, "Quality assessment for listening-room compensation algorithms," in *Proc. IEEE Int. Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Dallas, TX, USA, 2010.

[18] R. Huber and B. Kollmeier, "PEMO-Q—a new method for objective audio quality assessment using a model of auditory perception," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 14, no. 6, pp. 1902–1911, 2006.

[19] ITU-R Rec. BS.1284-1, *General methods for the subjective assessment of sound quality*, ITU, 2003.

[20] A. Raake, M. Wältermann, and S. Spors, "Which wideband speech codec? Quality impact due to room-acoustics at send side and presentation method," in *Proc. 127th AES Convention*, New York, NY, USA, 2009.

APPENDIX A: MEASUREMENT ROOMS

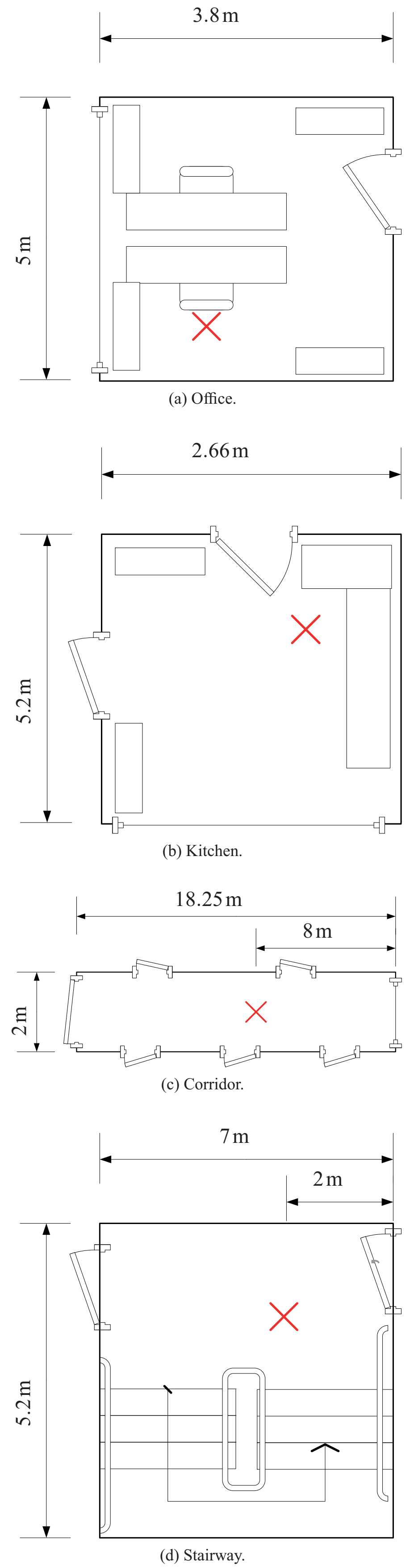


Figure 9: Room properties and measurement setup (not to scale). The cross marks the position of the dummy head.