Reproduction of Reverberation

Hans W. Gierlich, Magnus Schäfer, Torsten Ungerechts, Benedikt Koppers,
Jan Reimes, Shahd Al Hares, Christoph Nelke

HEAD acoustics GmbH, 52134 Herzogenrath, Deutschland, Email: telecom@head-acoustics.de

Abstract

Many devices that employ microphones to pick up speech signals are used in a hands-free manner. This includes, e.g., phones in hands-free mode, group-audio terminals or smart speakers with speech recognition capabilities like Amazon Echo or Google Home. Since the distance between talker and device may range from centimeters up to meters, the microphone signals contain a significant amount of noise and reverberation.

Testing of these devices requires a realistic reproduction of both the noise as well as the reverberation in a defined and reproducible manner. A background noise reproduction system is already available in ETSI TS 103 224 since 2014. An accompanying system for the reproduction of reverberation is described here that conforms to the recently approved specification ETSI TS 103 557.

This contribution presents the reproduction setup along with the necessary digital signal processing for a realistic reproduction of reverberant sound fields in a measurement chamber. A head and torso simulator is utilized in conjunction with an equalized eight-loudspeaker arrangement to add reverberation to arbitrary speech signals. The capabilities of the reproduction system are illustrated by a comparison of room acoustic parameters and measurements that were done with commercially available devices both in a reverberant room and in the corresponding reproduction.

Introduction

Reverberation is an important aspect of the acoustical environment. It is particularly critical for devices that are used in a hands-free manner like conference phones or smart speakers. Depending on the specific situation, it even has to be considered for hand-held telephony [1].

The other main interference that is usually encountered in real acoustic environments is background noise. A corresponding reproduction system was described and specified in ETSI TS 103 224 [2]. The original system was approved in August of 2014 with later extensions and modifications in August 2015 and July 2017. The reproduction system can be separated into two parts:

- Recording of sound fields in real acoustic environments
- Reproducing the sound fields in a test room

The recording is done with a microphone array. A fixed configuration is described in [2] that is particularly well suited for recording sound fields in the vicinity of an artificial head – focusing on the spatial area where a hand-held telephone is usually positioned, i.e., on the right side of the head, between mouth and ear. One of the aforementioned extensions made the reproduction system usable with arbitrary microphone arrangements. The most common setup features eight microphones and eight loudspeakers.

The reproduction stage consists of an equalized loudspeaker array. The equalization procedure requires using the same microphone array that was used for recording the sound fields in the test room as well. The procedure aims to invert the transmission paths from all loudspeakers to all equalization microphones in the test room. This allows to reproduce the recorded sound fields with a high degree of accuracy. A set of background noise signals that were recorded with the fixed microphone arrangement accompanies the specification [2]. This makes the test results when using the system comparable between different test facilities.

Due to the increasing popularity of smart speakers with voice recognition capabilities, it became apparent that the reproduction of background noise alone is not sufficient for a realistic evaluation of the device performance. Everyday experience shows that reverberation both impacts the communication quality and decreases the recognition rate. Accordingly, a work item was initiated in ETSI STQ to investigate the possibility of extending the reproduction system to reverberation. This programme lead to the specification ETSI TS 103 557 [3] that was approved recently in December of 2018.

A main difference to the reproduction system for background noise is directly apparent in the specification: Instead of relying on measured sound fields, the reverberation if characterized by the impulse responses $h_1 \dots h_M$ between the artificial mouth of a head and torso simulator (HATS) and the M microphones.

This contribution briefly introduces the reproduction system along with the necessary signal processing before investigating the reproduction accuracy of the system by reproducing seven very different reverberant rooms. The actual use case of the system is then illustrated by measuring the send frequency response and two instrumental measures for five different commercially available devices – once in a reverberant room and in the reproduction of that room.

System Overview

The signal processing of the reproduction system for reverberation is depicted in Figure 1, more details on the individual components can be found in [3]. The system uses two separate playback paths:

- The artificial mouth of a HATS plays the unfiltered source signal to reproduce the direct path of the impulse response.
- An equalized multi channel loudspeaker system that conforms to the requirements of ETSI TS 103 224 reproduces all reverberant components of the impulse response.

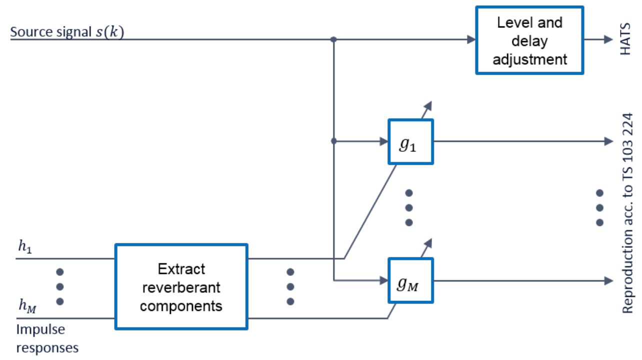

Figure 1: Block diagram of the signal processing for the reproduction of reverberation

Reproduction Setup

The reproduction setup requires a test room that fulfils the requirements of both [2] and [3]. The reproduction of reverberation recommends using an anechoic room – which usually inherently fulfils all the requirements for the background noise reproduction.

An example setup is shown in Figure 2. It consists of an artificial head that is positioned in the group-audio terminal position according to [4] in front of the device under test. Only three of the loudspeakers are visible. Together with the other five, they are positioned roughly equiangularly around the device under test at slightly varying heights and distances.

Room Impulse Response Measurements

To evaluate the accuracy of the reproduction system, room impulse responses were measured in seven different rooms. In accordance to the description in [3], three parameters were calculated from the measured impulse responses: Reverberation Time (RT_{60}), Clarity (C_{50}) and Direct-to-Reverberant Energy Ratio (DRR).

The reverberation time in this context shall not be confused with the purely room acoustical property as an artificial mouth is not the ideal sound source for measuring the room. It is, however, the ideal sound source for this application scenario of a talker addressing a device in a reverberant environment. The parameter RT_{60} is calculated from the measured impulse responses by the method described in [5] and a tolerance window of ± 50 ms is allowed between the value determined in the original room and the reproduced sound field.

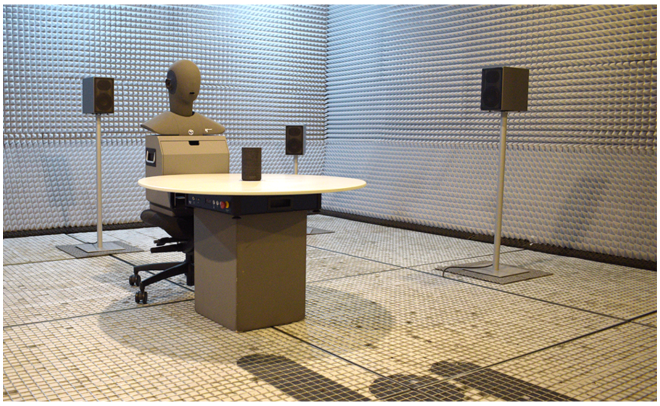

Figure 2: Example setup for the reproduction of reverberation

The other two parameters (C_{50} and DRR) are very similar as they both quantify the temporal structure of the impulse response by comparing the energy in early parts to the energy in the remainder of the impulse response. Both parameters are calculated from the impulse response $h(k)$ by (with f_S as the sampling frequency and K as the total length of the impulse response):

$$C_{t_{CO}} = 10 \cdot \log_{10} \left(\frac{\sum\limits_{0}^{t_{CO} \cdot f_S - 1} h^2(k)}{\sum\limits_{t_{CO} \cdot f_S}^{K} h^2(k)} \right) \tag{1}$$

The only difference is the summation limit t_{CO} which is 50 ms for C_{50} and 5 ms for DRR. The tolerance windows for the two values are ± 0.5 dB for C_{50} and ± 1.5 dB for DRR.

An overview of the dimensions and the measured parameters for the seven room is given in Table 1. It can be seen that the rooms are very different with respect to their properties. Thus, reproducing the acoustic properties of these rooms in a single test room is a fairly challenging task and the comparison between the original parameters and the parameters in the reproduction should uncover the capabilities and limitations of the reproduction system for most (if not all) realistic environments.

Reproduction Accuracy

The accuracy of the reproduction is quantified by a comparison between the room acoustical parameters as they were measured in the original rooms and those measured in the reproduced room. For the sake of brevity, the comparison in this publication omits the results for C_{50} as this parameter conceptually very similar to DRR. The results for RT_{60} are shown in Figure 3 with the values for the original rooms in blue and the values for the reproduction in green.

It is obvious that there are no major deviations between original and reproduction. From the shortest to the longest RT_{60}, all the values are reproduced very accurately. The average difference is 4 ms and even the maxi-

Table 1: Room dimensions and room acoustic parameters

Room	Base area [m^2]	Volume [m^3]	RT_{60} [ms]	C_{50} [dB]	DRR [dB]
Livingroom	42	105	388	13.4	4.3
Officeroom	19	58	544	10.7	2.0
Kitchen	9	23	547	7.8	0.0
Bathroom	12	24	583	9.3	2.9
medium Hall	28	172	1228	8.5	5.3
big Hall	200	1000	1847	12.3	10.4
Staircase	20	-	2277	4.6	0.5

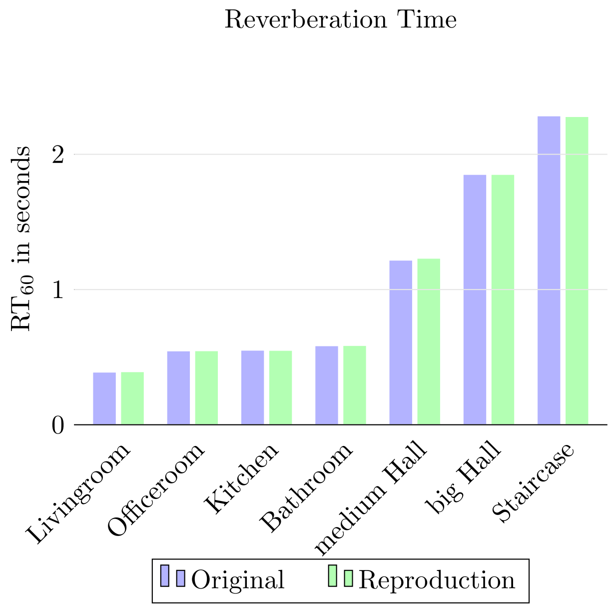

Figure 3: Comparison of reverberation times in reverberant and reproduced rooms

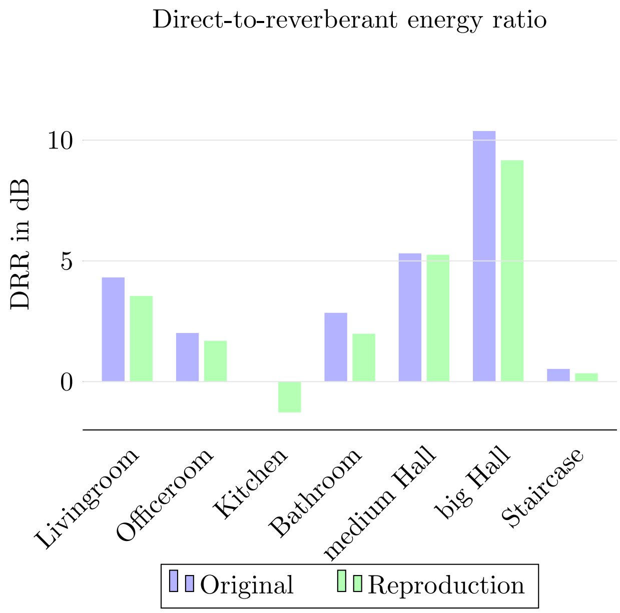

Figure 4: Comparison of direct-to-reverberant energy ratios in reverberant and reproduced rooms

mum difference is only 14 ms which is clearly smaller than the allowed tolerance.

The DRR values for the seven rooms are depicted in Figure 4. Again, the values for the original rooms are shown in blue, the values for the reproduction in green.

In contrast to the results for RT_{60}, there are visible differences. There are differences of more than 1 dB for the kitchen and the big hall in particular. The tolerance window is not violated, though. The average difference is 0.67 dB and the maximum difference occurs for the kitchen with 1.27 dB.

One point worth noting is that all values for DRR are smaller in the reproduction. Since the reproduction path for the direct sound is disjunct from the path for the remainder of the impulse response (cf. Figure 1), it would even be possible to increase the level of the direct sound to improve the accuracy for this parameter if needed.

Device Behaviour

It was shown in the previous section that the system is capable of reproducing room acoustic parameters to a high degree. The real application scenario considers a device in a reverberant sound field. Accordingly, a comparison of device behaviour in original and reproduced rooms is the next logical step to evaluate the system performance.

Five different devices were tested for this comparison. An overview of the device properties is given in Table 2. The device labels indicate the form factor of the device: S1 is small, M1 and M2 are medium, L1 and L2 are large devices. All devices were measured once in a reverberant room (Room1 from Annex A of [3]) and in the reproduction of that room.

Table 2: Properties of communication devices

Device	Characteristics
S1	Single directional microphone, single user
M1	Approximately omnidirectional, group of users
M2	Approximately omnidirectional, group of users
L1	Wide aperture beamformer, highly directive, single user
L2	Adaptive multi-microphone signal processing, group of users

The first parameter that is compared is the send frequency response according to [6]. The result for device M1 is shown in Figure 5. The red line is the frequency response as measured in the original room and the blue line is measured in the reproduction.

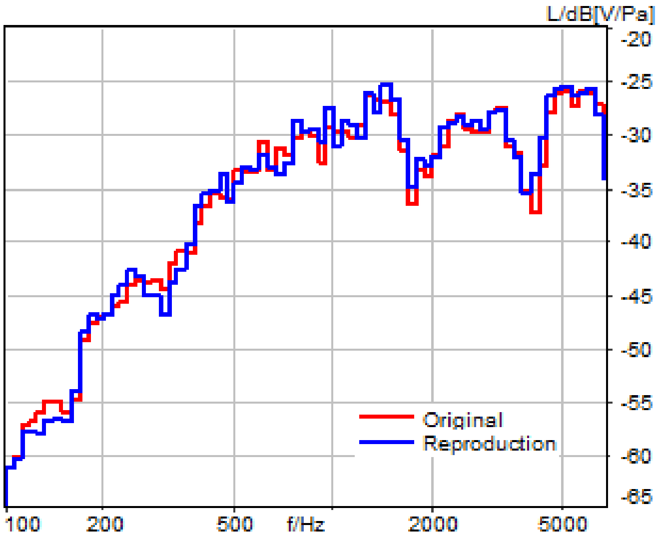

Figure 5: Send frequency responses of device M1 in original and reproduced room

It can be observed that there are small differences in single third octave bands but the overall shape is identical and there is no systematic error visible in the results. The same observation holds for all five devices.

The final comparison considers two different instrumental measures: ITU-T P.863 [7] and ETSI TS 103 281 [8]. Neither of these measures explicitly quantifies the impact of reverberation. Thus, the absolute values are not relevant for the comparison. Once again, the difference between the results in the original room and in the reproduction shall be as small as possible. The differences for ITU-T P.863 are in Table 3 while those for ETSI TS 103 281 are in Table 4.

Table 3: ITU-T P.863 – Difference between original and reproduction

Device	Original-Reproduction
S1	-0.027
M1	0.035
M2	-0.074
L1	-0.084
L2	0.039

Both instrumental measures provide values on mean opinion score (MOS) scales ranging from 1 to 5. The differences that are present are mostly negligible. Only device S1 has differences of more than 0.1 for N-MOS and G-MOS according to [8]. This device is a lot smaller than the microphone arrangement that was used. Thus, some differences can be expected and another microphone setup could be utilized if the sizes of device and microphone setup do not match.

Table 4: ETSI TS 103 281 – Difference between original and reproduction

Device	Original-Reproduction		
	S-MOS	N-MOS	G-MOS
S1	0.080	0.108	0.182
M1	-0.027	-0.035	-0.088
M2	0.047	0.006	0.017
L1	-0.085	0.033	0.030
L2	0.039	-0.015	0.024

Conclusions

An overview of a system for the reproduction of reverberation according to the recently approved specification ETSI TS 103 557 was presented. The system comprises recording of impulse responses in reverberant rooms, signal processing for adding reverberation to arbitrary source signals and an interface to the equalized loudspeaker playback from ETSI TS 103 224.

It was shown by several measurements and comparisons that the system is capable of reproducing room acoustic parameters accurately and that the behaviour of commercially available devices in the reproduction is very similar to the behaviour in a real reverberant room.

References

[1] Marco Jeub, Magnus Schäfer, Hauke Krüger, Christoph Nelke, Christophe Beaugeant, and Peter Vary. Do we need dereverberation for hand-held telephony? In *International Congress on Acoustics (ICA)*, Sydney, Australia, 2010.

[2] ETSI TS 103 224 V1.3.1. *A sound field reproduction method for terminal testing including a background noise database*, July 2017.

[3] ETSI TS 103 557 V1.1.1. *Methods for reproducing reverberation for communication device measurements*, December 2018.

[4] ITU-T Recommendation P.341. *Transmission characteristics for wideband digital loudspeaking and hands-free telephony terminals*, March 2011.

[5] Ning Xiang. Evaluation of reverberation times using a nonlinear regression approach. *The Journal of the Acoustical Society of America*, 98(4):2112–2121, 1995.

[6] ETSI TS 103 740 V1.3.1. *Transmission requirements for wideband wireless terminals (handsfree) from a QoS perspective as perceived by the user*, October 2018.

[7] ITU-T Recommendation P.863. *Methods for objective and subjective assessment of speech quality*, Sep. 2014.

[8] ETSI TS 103 281 V1.2.1. *Speech quality in the presence of background noise: Objective test methods for super-wideband and fullband terminals*, January 2018.