Available online at www.sciencedirect.com

SciVerse ScienceDirect

Speech Communication 54 (2012) 393–401

SPEECH COMMUNICATION

www.elsevier.com/locate/specom

On the quality-assessment of reverberated speech

Amaro A. de Lima [a,b,*], Thiago de M. Prego [b], Sergio L. Netto [b], Bowon Lee [c], Amir Said [c], Ronald W. Schafer [c], Ton Kalker [c], Majid Fozunbal [c]

[a] Federal Center for Technological Education Celso Suckow da Fonseca (CEFET-RJ), Nova Iguaçu, RJ, Brazil
[b] Program of Electrical Engineering, COPPE, Federal University of Rio de Janeiro, Rio de Janeiro, RJ, Brazil
[c] Hewlett-Packard Laboratories, 1501 Page Mill Road, Palo Alto, California 94304, USA

Received 18 December 2010; received in revised form 6 October 2011; accepted 7 October 2011
Available online 20 October 2011

Abstract

This paper addresses the problem of quantifying the reverberation effect in speech signals. The perception of reverberation is assessed based on a new measure combining the characteristics of reverberation time, room spectral variance, and direct-to-reverberant energy ratio, which are estimated from the associated room impulse response (RIR). The practical aspects behind a robust RIR estimation are underlined, allowing an effective feature extraction for reverberation evaluation. The resulting objective metric achieves a correlation factor of about 90% with the subjective scores of two distinct speech databases, illustrating the system's ability to assess the reverberation effect in a reliable manner.
© 2011 Elsevier B.V. All rights reserved.

Keywords: Speech quality evaluation; Reverberation assessment; Intrusive approach

1. Introduction

High-quality transmission systems are ever more present in human life in the form of HDTV, home-theater, professional teleconference/telepresence systems, and so on, requiring high-rate transfers of audio, video, and data signals. These top-notch systems must attain high levels of user satisfaction to deliver a realistic multimedia experience. Therefore, practical systems often incorporate quality-assessing tools to evaluate their performance in a reliable manner.

The three main acoustic impairments for speech transmissions between rooms A (source) and B (destination) are: background noise (possibly generated by an air-conditioner, a computer, or any other source in rooms A or B); echo (signal returns to room A through speaker-microphone coupling in room B); and reverberation (acoustical properties of rooms A and B are imposed on the signal).

This paper addresses the problem of estimating human perception of the reverberation effect on speech signals. One of the first attempts in this direction was described in a half-page abstract by Allen (1982), where a closed-form measure is proposed, and later validated in (Berkley and Allen, 1993). Related previous work includes references (Wen and Naylor, 2006; Wen et al., 2006), where the authors present the MARDY database, containing 32 speech signals with different levels of reverberation, and another quality evaluator. In (Falk et al., 2010), an objective-quality measure is determined directly from the reverberated signal alone, which is commonly referred to as a non-intrusive approach. More recently, (de Lima et al., 2009; Goetze et al., 2010) investigated the use of several individual metrics for assessing the reverberation effect,

* Corresponding author at: Federal Center for Technological Education Celso Suckow da Fonseca (CEFET-RJ), Nova Iguaçu, RJ, Brazil. Tel.: +55 21 9959 7151; fax: +55 21 3770 0064.
E-mail addresses: amaro@lps.ufrj.br (A.A. de Lima), thprego@lps.ufrj.br (T. de M. Prego), sergioln@lps.ufrj.br (S.L. Netto), bowon.lee@hp.com (B. Lee), amir_said@hp.com (A. Said), ron.schafer@hp.com (R.W. Schafer).

0167-6393/$ - see front matter © 2011 Elsevier B.V. All rights reserved.
doi:10.1016/j.specom.2011.10.003

trying to identify the most significant ones in a perceptual sense.

The present paper introduces an intrusive approach, which requires both the clean and degraded signals, for quality assessment of reverberation based on a closed-form measure combining results previously presented in (Allen, 1982; de Lima et al., 2009). Allen's original metric is modified to incorporate the direct-to-reverberant energy ratio in its original formulation. The resulting intrusive scheme is suited for off-line testing scenarios, where one evaluates the acoustical characteristics of a system prior to its operation. For developing and validating the complete measuring system, which we call QAreverb, a large database was deployed including 204 speech signals, with different levels of reverberation, which were subjectively evaluated by 30 listeners using the mean opinion score (MOS) 1-to-5 scale.

To describe all these contributions, this paper is organized as follows: In Section 2, the reverberation process is characterized along with three individual measures associated with that impairment; Section 3 describes the proposed system for evaluating the reverberation effect based on a new metric combining the measures introduced in Section 2; Section 4 describes the development of a large database of speech signals corrupted by reverberation along with the subjective tests performed on these signals; Section 5 evaluates the performance of the proposed QAreverb system in predicting the human perception of reverberation for two distinct speech databases, for which 91% and 95% statistical correlations are achieved with the respective subjective scores; Section 6 concludes the paper by summarizing its main contributions.

2. Standard reverberation features

The reverberation process is often modeled as the result of a convolution of a given audio signal with the room impulse response (RIR), $h(t)$, representing the acoustical characteristics of a room. In practice, one considers two distinct portions of the RIR, as depicted in Fig. 1:

- Early reflections: comprised of several impulses with amplitudes typically following an exponential decay and containing most of the RIR energy. In this context, the first impulse refers to the direct-sound component, defining the time instant t_d, and has a normalized amplitude set to 1.
- Late reverberation: constitutes the remaining RIR portion and presents a diffusive nature with no significant isolated impulsive components.

There are several measures associated with the reverberation effect (de Lima et al., 2009; Goetze et al., 2010; Kuttruff, 2000; Kuttruff, 2007; Figueiredo and Iazzetta, 2005). Three of them, however, seem to be the most important for perception (Allen, 1982; de Lima et al., 2009; Griesinger, 2009) and, for that reason, are employed by the proposed system to be described below.

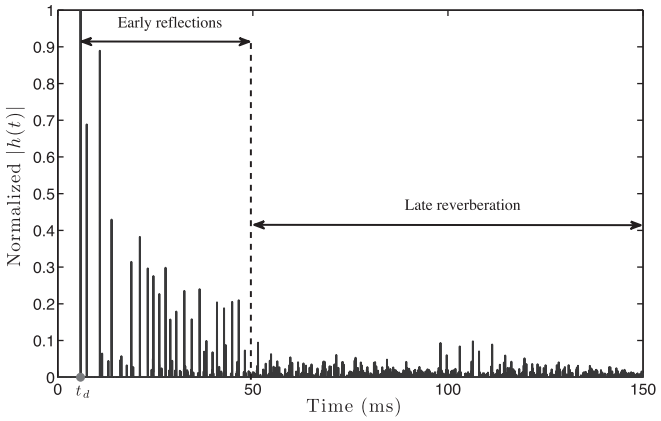

Fig. 1. RIR example with indication of the early reflection (with corresponding direct-sound arrival time t_d) and late reverberation portions.

2.1. Reverberation time

The reverberation time T_{60} is defined as the period of time required for the sound-pressure to decay 60 dB after the excitation signal (usually filtered noise) is turned off (Schroeder, 1965). In practice, a higher T_{60} indicates a more lasting reverberation effect. Measurement of T_{60} using noise excitation requires ensemble averaging over many trials, but Schroeder showed in Schroeder (1965) that T_{60} can be computed directly from a measured RIR. In particular, it was shown (Schroeder, 1965; Karjalainen et al., 2001) that the T_{60} of a given RIR, $h(t)$, can be estimated based on the normalized energy decay curve (EDC) defined as

$$\mathrm{EDC}(t) = 10\log_{10}\left(\frac{\int_t^\infty h^2(\tau)\mathrm{d}\tau}{\int_0^\infty h^2(\tau)\mathrm{d}\tau}\right)[\mathrm{dB}], \quad (1)$$

where the denominator guarantees a maximum EDC value of 0 dB at $t = 0$. On the dB scale, the EDC can be approximated by a first-order function (indicated, for example, by the dashed line in Fig. 2), usually starting at the -5 dB level and up to a stop point at which one considers the reverberated signal to be significantly affected by noise (see (Schroeder, 1965; Karjalainen et al., 2001) for different stop-point criteria). The estimated T_{60} may then be determined as the time interval required by this first-order EDC approximation to fall from 0 to -60 dB.

2.2. Room spectral variance

Let $H(f)$ be the Fourier transform of $h(t)$. The relative acoustic intensity level is defined as (Jetz, 1979)

$$I(f) = 10\log_{10}\left[\frac{|H(f)|^2}{\overline{|H(f)|^2}}\right][\mathrm{dB}], \quad (2)$$

where the overbar $\{\bar{\cdot}\}$ denotes the average of a function across all frequency values f. The room spectral variance

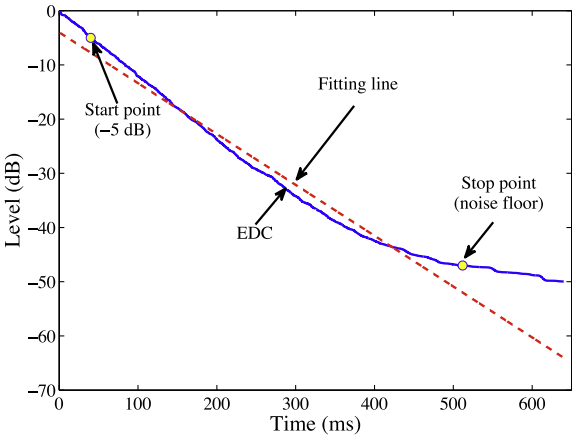

Fig. 2. Reverberation time estimation: The EDC curve (solid line) generates a linear fitting (dashed line), the slope of which is used to determine the time-interval T_{60} between the 0 and −60 dB levels. The stop point is commonly chosen based on a noise-floor criteria. In this case the estimated reverberation time is $T_{60} \approx \frac{(510-40)\ [\mathrm{ms}]}{(47-5)\ [\mathrm{dB}]} 60\ [\mathrm{dB}] \approx 670\ [\mathrm{ms}]$.

(RSV) is determined by the variance of $I(f)$ in dB in the frequency domain, that is

$$\sigma_I^2 = \overline{(I(f) - \overline{I(f)})^2}. \tag{3}$$

The RSV characterizes the reverberation effect in the frequency domain. In fact, a flatter magnitude response $|H(f)|$, which corresponds to low RSV values, is less perceived than a spiky response, which provides a coloration effect to a speech sound.

2.3. Direct-to-reverberant energy ratio

This feature is defined as the ratio between the direct E_d (within a short interval around t_d) and reverberant E_r (remaining) energy levels of $h(t)$, as given by Zahorik (2002a,b)

$$R = \frac{E_d}{E_r} = \frac{\int_{(t_d-1)\mathrm{ms}}^{(t_d+1.5)\mathrm{ms}} h^2(t)\mathrm{d}t}{\int_{(t_d+1.5)\mathrm{ms}}^{\infty} h^2(t)\mathrm{d}t}. \tag{4}$$

To reduce noise influence, one may consider only the signal components 20 dB above the noise floor level in $h(t)$ and halt the energy accumulation at the stop point employed by T_{60} algorithm, as suggested in (Kuster, 2008).

3. New reverberation assessment

The T_{60} is arguably the most important feature for quantifying the reverberation effect on a given audio sample. In (Allen, 1982), Allen proposed a reverberation measure combining the reverberation time and RSV features, which was later validated in (Berkley and Allen, 1993). According to Griesinger (2009), Cole et al. (1994), however, the direct-to-reverberant energy ratio R provides a fundamental cue to assess speech intelligibility in closed

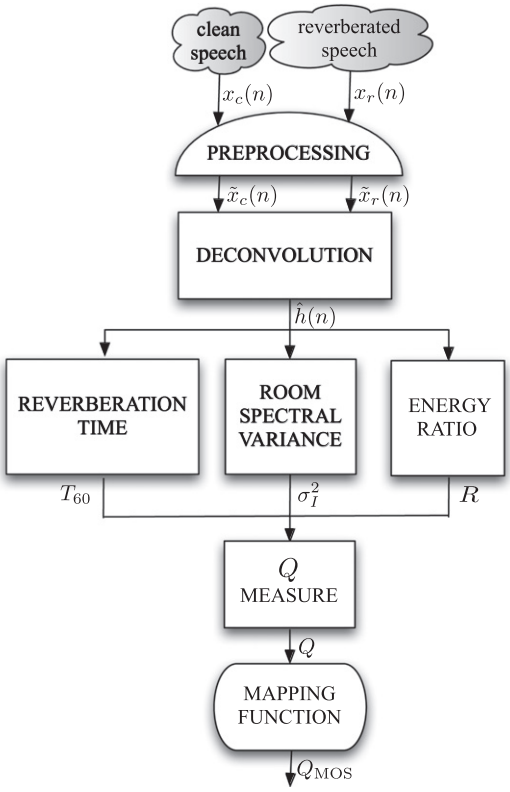

Fig. 3. Block diagram of QAreverb system based on new reverberation measure Q_{MOS}.

spaces, delivering to the listener some sense of source distance and localization. Practical experiments included in (de Lima et al., 2009) also indicate the importance of the direct-to-reverberant energy ratio on the subjective perception of reverberation. In fact, although the RSV feature is closely related to R when $R \geqslant 1$, these two metrics become disassociated for large source-microphone distances (Larsen et al., 2008) or, correspondingly, when $R < 1$ (Jetz, 1979).

Hence, a new measure Q for quality assessment of reverberation is proposed, combining the three features presented in Section 2, incorporating the energy ratio R into Allen's score as given by

$$Q = -\frac{T_{60}\sigma_I^2}{R^{\gamma}}, \tag{5}$$

with the exponent γ determined empirically in the system-training stage, where the special case $\gamma = 0$ corresponds to Allen's original measure (Allen, 1982).

The general reverberation-assessing system, based on this new measure, may be implemented in the discrete-time domain n for any sampling frequency F_s as depicted in Fig. 3. This system receives the clean $x_c(n)$ and reverberated $x_r(n)$ discrete-time versions of a given speech signal, which are employed to obtain an estimate $\hat{h}(n)$ of the associated RIR. The T_{60}, σ_I^2, and R measures are then estimated from $\hat{h}(n)$, allowing one to determine the reverberation score as

defined in Eq. (5). In this scheme, the role of each block in Fig. 3 is as follows:

- Preprocessing: removes the DC level from the clean $x_c(n)$ and reverberated $x_r(n)$ speech signals, generating the processed signals $\tilde{x}_c(n)$ and $\tilde{x}_r(n)$, respectively.
- Deconvolution: estimates the RIR $\hat{h}(n)$ by performing the deconvolution between $\tilde{x}_c(n)$ and $\tilde{x}_r(n)$, as given by

$$\hat{h}(n) = \mathrm{IDFT}\left[\frac{\mathrm{DFT}[\tilde{x}_r(n)]}{\mathrm{DFT}[\tilde{x}_c(n)]}\right], \tag{6}$$

where DFT[·] and IDFT[·] represent the discrete Fourier transform and its inverse operation, respectively. If we consider a 1-s RIR $\hat{h}(n)$, all (I)DFT operations in Eq. (6) can be performed by fast algorithms, such as the fast Fourier transform (FFT), of size given by the lowest power-of-2 greater than or equal to $(F_s + \ell_c - 1)$, where ℓ_c is the clean-signal length. When implementing Eq. (6), one must deal with the problem presented by small values of the denominator. Our approach is to impose a lower limit $\epsilon > 0$ on the absolute value of $\mathrm{DFT}[\tilde{x}_c(n)]$; that is, $\widetilde{X}_c(k) = \mathrm{DFT}[\tilde{x}_c(n)]$ is replaced by ϵ for all indices k such that $|\widetilde{X}_c(k)| < \epsilon$, or, more specifically,

$$\text{for all } k : \text{if} |\widetilde{X}_c(k)| < \epsilon \Rightarrow \widetilde{X}_c(k) \leftarrow \epsilon, \tag{7}$$

where $A \leftarrow B$ means A is replaced by B. The influence of this parameter ϵ on the proposed system performance is discussed later in Section 5.1.

- Reverberation time: estimates T_{60} from $\hat{h}(n)$ as described in Section 2.1. In this work, we have used the algorithm presented in (Karjalainen et al., 2001), as provided by its authors.
- Room spectral variance: determines σ_I^2 associated with the estimated RIR using a discrete-frequency version of Eq. (3). Hence, σ_I^2 is determined as the variance of $I(f)$, defined in Eq. (2), with $H(f)$ replaced by $H(k) = \mathrm{DFT}[\hat{h}(n)]$.
- Energy ratio: computes the direct-to-reverberant energy ratio R for the estimated RIR using the discrete-time counterpart of Eq. (4), where t_d is the time instant associated to the maximum value of $|\hat{h}(n)|$, that is, $t_d = \arg\{max_n |\hat{h}(n)|\}/F_s$.
- Proposed measure: determines the reverberation score Q as defined in Eq. (5).
- Mapping function: maps the values of Q onto the MOS scale, using a third-order polynomial of the form (ITU-T Rec. P.563, 2004)

$$\overline{Q} = x_1 Q^3 + x_2 Q^2 + x_3 Q + x_4, \tag{8}$$

where the coefficients x_1, x_2, x_3, and x_4 are determined during system training. In practice, different grades may be given by listeners when different reverberation ranges are considered in the subjective test. Therefore, this procedure may be followed by a linear-scale adjustment of $\overline{Q}$ to the grade scale of a distinct subjective test, as given by Zielinski and Rumsey (2008)

$$Q_{\mathrm{MOS}} = \alpha \overline{Q} + \beta, \tag{9}$$

with α and β possibly determined by some data subset. This linear mapping reduces the mean squared error (MSE) between objective and subjective scores without changing the associated correlation factor (Kay, 1993).

4. Databases of reverberant speech

Two distinct databases containing samples of reverberated speech were deployed in the development of the QAreverb system: the MARDY database (Wen et al., 2006) and a new Brazilian–Portuguese (NBP) database specifically developed in the present context.

4.1. MARDY database

The MARDY database (Wen et al., 2006) includes 16 reverberant, naturally degraded signals recorded directly in an auditorium, and their 16 dereverberated versions using delay-and-sum algorithm, making a total of 32 speech signals with sampling frequency $F_s = 16$ kHz. This database considers 2 different speakers (1 male and 1 female), 4 values for the source-microphone distance ($d = 1$, 2, 3, 4 m), and 2 types (reflective and absorbent) of wall panels, which correspond to an estimated T_{60} of 447 ms and 291 ms, respectively.

The MARDY database was probably the first one developed for reverberation-assessment purposes in speech. However, it contains only a small number of signals all recorded in a single room. These characteristics motivated the development of a larger and more general database, as described in the following subsection.

4.2. New database of reverberant speech

The NBP database was completely developed based on 4 anechoic signals uttered by 2 speakers (1 male and 1 female) with an $F_s = 48$ kHz sampling rate. Each signal was comprised of two short Brazilian–Portuguese sentences separated by approximately 1.7 s, giving an 8.4-s average duration for the entire database. The reverberation effect was imposed onto these anechoic signals using three distinct approaches, namely:

- Artificial reverberation: In this method, the reverberation effect was emulated by 6 artificially generated RIRs, giving a total of 24 signals. In these RIRs, the early reflections were modeled via the image method (Allen and Berkley, 1979), with a fixed source-microphone distance $d = 1.8$ m in a virtual room of dimensions length × width × height=4 m × 3 m × 3 m. As regards the late reverberation, the feedback delay network method (Jot and Chaigne, 1991) was used to emulate reverberation times in the range $T_{60} = 200$, 300, 400 ms and a modified version of Gardner's method

Table 1
Room characteristics for natural reverberation effect in NBP database.

Room Type	Dimensions [$m \times m \times m$]	T_{60} [ms]	d [m]
Booth	$3.0 \times 1.8 \times 2.2$	120	0.5, 1, 1.5
Office	$5.0 \times 6.4 \times 2.9$	430	1, 2, 3
Meeting	$8.0 \times 5.0 \times 3.1$	230	1.45, 1.7, 1.9, 2.25, 2.8
Lecture	$10.8 \times 10.9 \times 3.15$	780	2.25, 4, 5.6, 7.1, 8.7, 10.2

(Gardner, 1998; de Lima et al., 2008), which was originally devised to emulate reverberation times above 400 ms, was used for $T_{60} = 500$, 600, 700 ms.

- Natural reverberation: In this approach, 17 distinct RIRs, as provided in (Jeub et al., 2009), were used to convolve the 4 anechoic signals to generate a total of 68 reverberant signals. These RIRs were obtained from 4 different rooms and distinct source-microphone distances (d), as summarized in Table 1.
- Real reverberation: In this method, the 4 anechoic signals were played/recorded in 7 rooms with different reverberating characteristics. In each room, 4 values for the source-microphone distance d were considered, except in the smaller room where only 3 distances were employed, as detailed in Table 2, giving a total of 108 signals with real reverberation. The T_{60} values in this table are the average values for all distances d in each room. The high T_{60} values associated to rooms 'meeting2' and 'office2' result from the highly reflective characteristic of their walls.

Subjective tests were performed for the 204 NBP (24 artificial, 68 natural, 108 real, and 4 anechoic) signals using the absolute category rate (ACR) MOS test (ITU-T Rec. P.800, 1996) with 30 listeners per signal. An additional 10 signals, covering the entire NBP reverberation range, were used in the initial part of the test to assist the listener in adjusting his/her scoring scale. Without the listener's knowledge, these initial scores were discarded later on. In the end, outliers were removed by establishing a score range of three standard deviations around the mean score of each signal. Only 9, all from different listeners and signals, out of a total of 6120 scores were removed in this procedure. The MOS results along with the corresponding standard deviation for each NBP signal are depicted in Fig. 4, in increasing MOS order. Error-margin results are quantitatively comparable to ITU-T subjective scores employed in evaluating the PESQ algorithm (ITU-T Rec. P.862, 2001).

5. Reverberation assessment of speech signals

5.1. Choosing ϵ

Different recording setups yield different spectral shapes, as exemplified in Fig. 5, which shows plots of the DFT of one NBP and one MARDY signal. To aid in comparing the spectra, which have different sampling rates (MARDY, 16 kHz and NBP, 48 kHz), we have plotted both on a normalized frequency scale with a corresponding sampling frequency $F_s' = 2$.

These representative plots show that the anti-alias filter used in the MARDY database had a relatively wider transition region than that of the filter used for the NBP database. Therefore, the value of ϵ in Eq. (7) must be tuned for each database to reduce the numerical errors in the

Table 2
Room characteristics for real reverberation effect in NBP database.

Room type	Dimensions [$m \times m \times m$]	T_{60} [ms]	d [m]
Booth	$2.1 \times 1.8 \times 2.4$	140	0.5, 1, 1.5
Office1	$7.4 \times 5.0 \times 2.7$	390	1, 2, 3, 4
Lecture1	$15.0 \times 10.0 \times 4.0$	570	1, 2, 3, 4
Meeting1	$10.0 \times 4.8 \times 3.2$	650	1, 2, 3, 4
Lecture2	$16.5 \times 8.2 \times 3.5$	700	1, 2, 3, 4
Meeting2	$9.0 \times 7.3 \times 3.5$	890	1, 2, 3, 4
Office2	$7.4 \times 4.8 \times 4.3$	920	1, 2, 3, 4

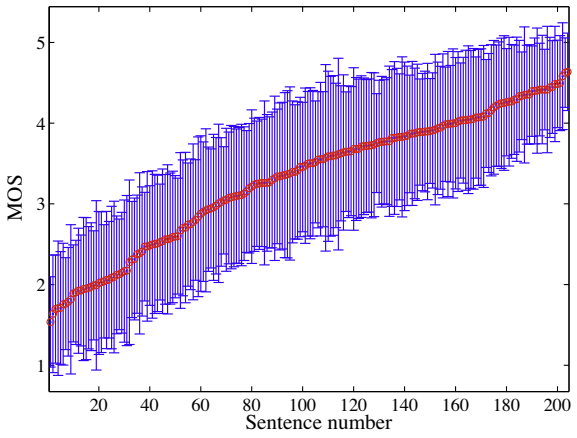

Fig. 4. Subjective MOS results and corresponding standard deviation for each NBP signal.

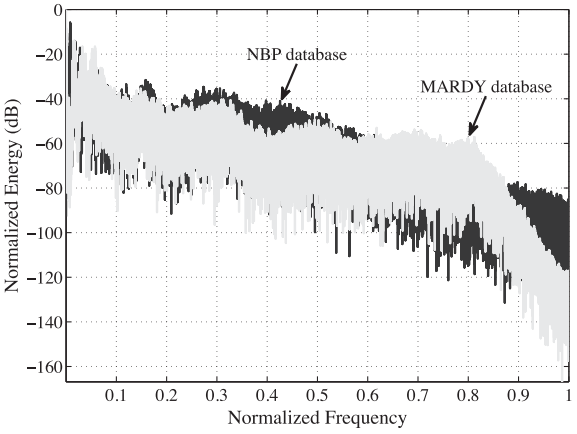

Fig. 5. Normalized spectral content of a clean speech signal of NBP (black) and MARDY (light gray) databases, as appears in the denominator of Eq. (6).

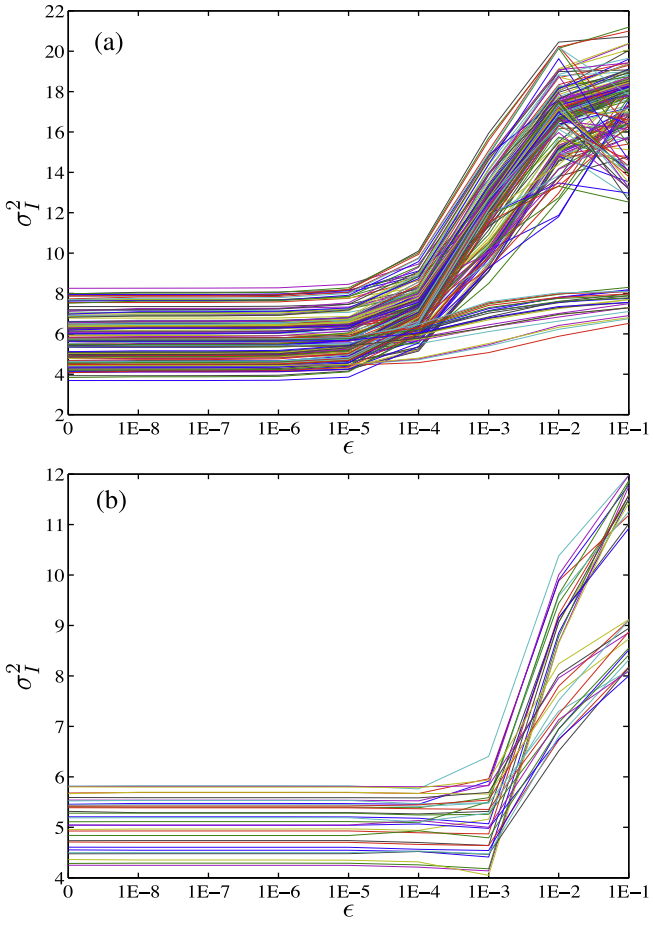

Fig. 6. Variation of σ_I^2 with respect to regulation parameter ϵ for: (a) All 200 non-anechoic NBP signals; (b) All 32 MARDY signals. These plots indicate that all (non-anechoic) signals in a given database present a similar RSV variation as a function of ϵ. This pattern, when determined for a particular signal, can be used to estimate a practical value of ϵ for the entire database.

associated RIR-estimation process. In practice, however, this procedure can be done in a very simple and robust manner by considering the behavior of σ_I^2, as a function of ϵ, as seen in Fig. 6 for both databases. These plots indicate that small values of ϵ lead to a similar flat behavior of σ_I^2 for all signals in each database, up to a threshold value above which σ_I^2 changes significantly. The increasing RSV patterns shown in Fig. 6 are explained by the fact that large values of epsilon force $|H(f)|$ to become closer to 0, which corresponds to $-\infty$ in the dB scale, thus increasing the resulting value of σ_I^2 as determined by Eqs. (2) and (3). As small values of ϵ lead to numerical errors in Eq. (6), affecting other reverberation aspects, a good strategy is to select ϵ within the flat RSV portion and close to the threshold value. This analysis can be performed for any particular signal and then extended to the entire database, leading to the values of $\epsilon = 10^{-5}$ and $\epsilon = 10^{-3}$ selected for the NBP and MARDY databases, respectively. The much larger ϵ required by the MARDY database may be explained by the poorer spectral distribution depicted in Fig. 5, particularly in the normalized-frequency range $0.9 \leqslant f' \leqslant 1$.

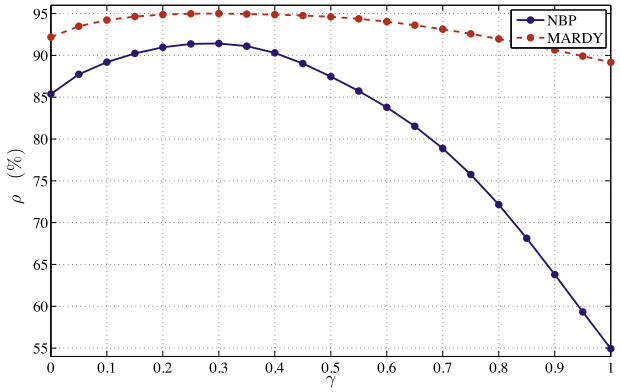

Fig. 7. Influence of energy-ratio exponent γ on the correlation factor ρ between objective Q_{MOS} score and subjective MOS for the NBP and MARDY databases.

5.2. QAreverb with the NBP database

Using $\epsilon = 10^{-5}$, a proper choice for the parameter γ in the definition of Q given in Eq. (5) was determined by measuring the correlation coefficient ρ between the resulting score and the subjective grades for the NBP database. As can be observed in Fig. 7, for the NBP database, the value of $\gamma = 0.3$ yields the maximum correlation score $\rho = 91\%$, which requires the nonlinear-mapping coefficients $x_1 = 0.0017$, $x_2 = 0.0598$, $x_3 = 0.7014$, and $x_4 = 4.5387$. in Eq. (8), as obtained by numerical optimization. In this figure, for fairness purposes, an optimal mapping (x_1, x_2, x_3, x_4) was determined for each value of γ, including the specific cases of $\gamma = 0$ and $\gamma = 0.3$. In order to minimize the MSE between Q_{MOS} (with $\gamma = 0.3$) and the NBP subjective scores, the coefficients $\alpha = 1.0000$ and $\beta = 1.85 \times 10^{-10}$ are used in Eq. (9), leading to the QAreverb results shown in Fig. 8.

For the "Artificial" and "Natural" NBP partitions, the Q_{MOS} measure was determined by estimating the T_{60}, RSV σ_I^2, and direct-to-reverberant energy ratio R from two distinct RIR versions: (i) the "true" RIR; (ii) the estimated RIR, as determined by Eq. (6). For both values of $\gamma = 0$ and $\gamma = 0.3$, the correlation between the two resulting Q_{MOS} scores was 99.9%, indicating that the estimated RIR, as provided by Eq. (6) with a proper choice of ϵ, does not cause any significant impact on the QAreverb performance.

Table 3 shows the statistical correlation between the subjective scores for all NBP subdivisions and the objective results by several speech-evaluation algorithms, such as (signal bandwidth is specified in parentheses): ITU W-PESQ (7 kHz) (ITU-T Rec. P.862.2, 2005), ITU P.563 (4 kHz) (ITU-T Rec. P.563, 2004), the reverberation decay time (R_{DT}, 4 kHz) (Wen and Naylor, 2006; Wen et al., 2006), speech-to-reverberation modulation energy ratio (SRMR, 4 or 8 kHz) (Falk et al., 2010), and Q_{MOS} measure for $\gamma = 0$ and $\gamma = 0.3$ (any bandwidth up to 24 kHz). In this set, the P.563 and SRMR constitute non-intrusive algorithms that depend only on the reverberated speech signal, whereas the other schemes require also the clean signal, for

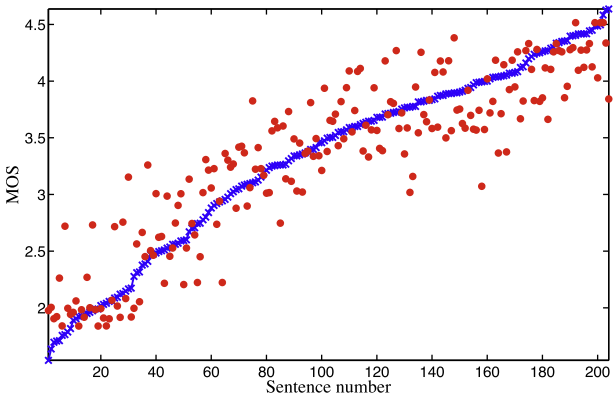

Fig. 8. Reverberation-assessment scores for all 204 sentences of NBP database: MOS (connected ' × ') and Q_{MOS} score with $\gamma = 0.3$ (scattered '•').

Table 3
Statistical correlation ρ (without/with optimal third-order mapping described in Eq. (8)) between subjective grades and objective scores by several quality-evaluating algorithms for the NBP database.

Objective algorithm	Correlation (ρ)[%]			
	Artificial	Natural	Real	Entire NBP
W-PESQ	84/72	84/94	86/93	77/89
P.563	08/14	64/65	45/60	52/59
R_{DT}	68/69	75/80	43/43	59/61
SRMR	73/77	80/84	70/80	74/81
Q_{MOS} ($\gamma = 0$)	90/89	92/92	86/88	85/85
Q_{MOS} ($\gamma = 0.3$)	90/91	85/96	80/88	81/91

example, to generate the RIR estimate as in the R_{DT} and Q_{MOS} cases. In general, intrusive methods tend to perform better than the non-intrusive ones, which operate blindly by processing only the corrupted version of the speech. The impact of signal bandwidth on reverberation assessment is still an open issue in the associated literature. In any case, both databases were properly downsampled to comply with the signal bandwidth (8-kHz mode for the SRMR) of each objective evaluator. Table 3 includes correlation results for all algorithms without/with the third-order mapping described in Eq. (8), using an optimal set of coefficients (x_1, x_2, x_3, x_4) for each algorithm. Each mapping was designed by maximizing the algorithm's correlation score for the entire NBP database, which explains any correlation decrease in the first three columns of Table 3 when using the mapping. Despite the fact that the ITU standards were not originally conceived for reverberation assessment, the W-PESQ algorithm presented a surprisingly good correlation level, especially when incorporating the optimized third-order mapping. For the non-intrusive case, the SRMR represents the current state-of-the-art, achieving 81% correlation. From Table 3, one concludes that the QAreverb system, when using the third-order mapping and $\gamma = 0.3$, was able to outperform all other methods, including the particular case of $\gamma = 0$ which corresponds to Allen's original score, achieving the highest correlation level of 91% for the complete NBP database.

Using Fisher's z-test to compare the Q_{MOS} correlations for $\gamma = 0$ and $\gamma = 0.3$ in the NBP database, one obtains $p = 0.015$, which provides a 98.5% confidence level that they correspond to statistically different distributions.

5.3. QAreverb with the MARDY database

The MARDY database is used here to validate the QAreverb performance, as it contains reverberant speech signals not employed in system design. Referring back to Fig. 7, one observes that the values of $\gamma = 0.3$, $x_1 = 0.0017$, $x_2 = 0.0598$, $x_3 = 0.7014$, and $x_4 = 4.5387$, adjusted for the NBP database also provide a very high correlation score for the MARDY database, which in this case rises up to $\rho = 95\%$. This higher correlation level, as compared to the NBP 91% score, may be explained by the narrower bandwidth and smaller T_{60} range considered by the MARDY database, which lead to a simpler and more easily modeled process.

As detailed in (Wen and Naylor, 2006; Wen et al., 2006), the MARDY listening test is composed of three different experiments concerning the subjective perception of coloration, reverberation tail effect, and overall speech quality, all using the MOS scale. In this case, the QAreverb system computes only one score and Table 4 shows the statistical correlation between this value (or the ones from the same speech-evaluation algorithms considered in the previous subsection, once again without/with the third-order mapping optimized for the NBP database) with the subjective scores for all three MARDY reverberation aspects. A breakdown is also provided in Table 4 for the reverberant

Table 4
Statistical correlation ρ (without/with optimal third-order mapping described in Eq. (8)) between subjective grades and objective scores by several quality-evaluating algorithms for the MARDY database.

Objective algorithm	Correlation (ρ) [%]				
	Coloration	Tail effect	Reverberant	Delay-and-Sum dereverberated	Overall reverberation
W-PESQ	70/72	80/87	69/75	78/81	72/77
P.563	44/46	49/51	61/59	42/42	55/54
R_{DT}	59/61	59/60	70/69	51/52	64/64
SRMR	84/76	82/82	78/78	80/74	79/77
Q_{MOS} ($\gamma = 0$)	90/90	97/97	91/91	93/93	91/92
Q_{MOS} ($\gamma = 0.3$)	88/90	94/95	97/96	94/95	95/95

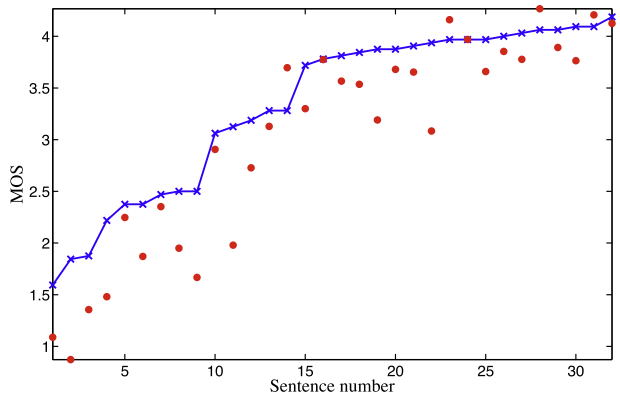

Fig. 9. Reverberation-assessment scores for all 32 sentences of MARDY database: MOS (connected ' × ') and Q_{MOS} score with $\gamma = 0.3$ (scattered '•').

and dereverberated portions of the MARDY database. From these results, once again the QAreverb system with $\gamma = 0.3$ was able to outperform all other methods, yielding the highest correlation level for the overall-reverberation scope, which is the main focus of the proposed evaluation measure.

Since the subjective tests performed for both databases comprised different ranges of degradation, another set of coefficients $\alpha = 1.3314$ and $\beta = -1.4224$ is required in Eq. (9) to adjust the QAreverb system for the MARDY database, as discussed in the end of Section 3, leading to the results shown in Fig. 9.

The statistical difference for the Q_{MOS} correlations with $\gamma = 0$ and $\gamma = 0.3$ for the MARDY database was established using Fisher's z-test, yielding $p = 0.30$, which corresponds to a 70% confidence level on the hypothesis of distinct distributions. The lower confidence level in this case, as compared to its NBP counterpart, can be attributed to the smaller number of signals, 32 as opposed to 204, contained in the MARDY database.

6. Conclusion

This paper addressed the task of estimating the perceived effect of reverberation on speech signals. A complete quality-evaluation system was described based on a modified Allen's score by incorporating the direct-to-reverberant energy ratio. An entire system-tuning procedure was detailed and practical validation which was provided using a new database, comprising a total of 204 reverberated signals, which was developed including a subjective evaluation from 30 listeners for each signal. The system was tested using two independent databases, leading to correlation scores of 91% and 95% with the corresponding subjective evaluations, improving on Allen's original proposal and outperforming other reverberation-assessing methods found in the literature.

Acknowledgments

The authors would like to thank Prof. M. Karjalainen, for providing the T_{60} estimation algorithm; Dr. J. Y. C. Wen, for making the MARDY database available for this research along with the R_{DT} routine; and Dr. T. H. Falk, for providing the SRMR software.

References

Allen, J.B., 1982. Effects of small room reverberation on subjective preference. J. Acoustic. Soc. Am. 71.

Allen, J.B., Berkley, D.A., 1979. Image method for efficiently simulating small-room acoustics. J. Acoustic. Soc. Am. 65 (4), 943–950.

Berkley, D.A., Allen, J.B., 1993. Normal listening in typical rooms: the physical and psychophysical correlates of reverberation. In: Studebaker, G.A., Hochberg, I. (Eds.), Acoustical Factors Affecting Hearing Aid Performance, 2nd ed. Allyn and Bacon.

Cole, D., Moody, M., Sridharan, S. 1994. Intelligibility of reverberant speech enhanced by inversion of room response. In: Proc. Int. Symp. on Speech, Image Processing, and Neural Networks, Hong Kong, pp. 241–244.

de Lima, A.A., Freeland, F.P., Esquef, P.A.A., Biscainho, L.W.P., Bispo, B.C., de Jesus, R.A., Netto, S.L., Schafer, R., Said, A., Lee, B., Kalker, A. 2008. Reverberation assessment in audioband speech signals for telepresence systems. In: Proc. Int. Conf. Signal Processing in Multimedia Applications, Porto, Portugal, pp. 257–262.

de Lima, A.A., de, T., Prego, M., Netto, S.L., Lee, B., Said, A., Schafer, R.W., Kalker, T., Fozunbal, M. 2009. Feature analysis for quality assessment of reverberated speech. In: Proc. Int. Workshop Multimedia Signal Processing, Rio de Janeiro, Brazil.

Falk, T.H., Zheng, C., Chan, W.-Y., 2010. A non-intrusive quality and intelligibility measure of reverberant and dereverberated speech. IEEE Trans. Audio Speech Lang. Process. 18 (7).

Figueiredo, F.L., Iazzetta, F. 2005. Comparative study of measured acoustic parameters in concert halls in the city of São Paulo. In: Proc. Int. Congress and Exposition on Noise Control Engineering, Rio de Janeiro, Brazil

Gardner, W.G., 1998. Reverberation Algorithms. In: Kahrs, Mark, Brandenburg, Karl-Heinz (Eds.), Applications of Digital Signal Processing. Kluwer, New York: NY, pp. 85–131.

Goetze, S., Albertin, E., Kallinger, M., Mertins, A., Kammeyer, K.-D. 2010. Quality assessment for listening-room compensation algorithms. In: Proc. IEEE Int. Conf. on Acoustics, Speech, and Signal Processing Dallas, USA.

Griesinger, D. 2009. The importance of the direct to reverberant ratio in the perception of distance, localization, clarity, and envelopment, Parts 1 and 2," In: 157th Meeting Acoustic. Soc. Am., Portland, USA.

ITU-T Rec. P.563, Single-ended Method for Objective Speech Quality Assessment in Narrow-band Telephony Applications, 2004.

ITU-T Rec. P.800, Methods for Subjective Determination of Transmission Quality, 1996.

ITU-T Rec. P.862, Perceptual evaluation of speech quality (PESQ): an objective method for end-to-end speech quality assessment of narrow-band telephone networks and speech codecs, 2001.

ITU-T Rec. P.862.2, Wideband extention to recommendation P.862 for the assessment of wideband telephone networks and speech codecs, 2005.

Jetz, J.J., 1979. Critical distance measurement of rooms from the sound energy spectral response. J. Acoustic. Soc. Am. 65, 1204–1211.

Jeub, M., Schäfer, M.,Vary, P. 2009. A binaural room impulse response database for the evaluation of dereverberation algorithms. In: Proc. 16th Int. Conf. on Digital Signal Processing, Santorini, Greece.

Jot, J.-M., Chaigne, A. 1991. Digital delay networks for designing artificial reverberators. In: Proc. 90th Conv. Am. Engineering Soc., Preprint 3030.

Karjalainen, M., Antsalo, P., Mäkivirta, A., Peltonen, T., Välimäki, V. 2001. Estimation of modal decay parameters from noisy reponse measurements. In: Proc. Conv. Audio Engineering Society, Amsterdam, Netherlands, pp. 867–878.

Kay, S.M., 1993. Fundamentals of Statistical Signal Processing: Estimation Theory. Prenctice-Hall, Upper Saddle River:NJ.

Kuster, M., 2008. Reliability of estimating the room volume from a single room impulse response. J. Acoustic. Soc. Am. 124, 982–993.

Kuttruff, H., 2000. Room Acoustics, 4th ed. Taylor & Francis, New York, USA.

Kuttruff, H., 2007. Acoustics, An Introduction. Taylor & Francis, New York, USA.

Larsen, E., Iyer, N., Lansing, C.R., Feng, A.S., 2008. On the minimum audible difference in direct-to-reverberant energy ratio. J. Acoustic. Soc. Am. 124, 450–461.

Schroeder, M.R., 1965. New method of measuring reverberation time. J. Acoustic. Soc. Am. 37 (3), 409–412.

Wen, J.Y.C., Naylor, P.A. 2006. An evaluation measure for reverberant speech using tail decay modeling. in: Proc. European Signal Processing Conf., Florence, Italy.

Wen, J.Y.C., Gaubitch, N.D., Habets, E.A.P., Myatt, T., Naylor, P.A. 2006. Evaluation of speech dereverberation algorithms using the MARDY database. in: Proc. IEEE Int. Workshop Acoustic Echo and Noise Control, Paris, France.

Zahorik, P., 2002a. Assessing auditory distance perception using virtual acoustics. J. Acoustic. Soc. Am. 111, 1832–1846.

Zahorik, P., 2002b. Direct-to-reverberant energy ratio sensitivity. J. Acoustic. Soc. Am. 112, 2110–2117.

Zielinski, S., Rumsey, F., 2008. On some biases encountered in modern audio quality listening test - a review. J. Audio Eng. Soc. 56 (6), 427–451.