EVALUATION OF SPEECH DEREVERBERATION ALGORITHMS USING THE MARDY DATABASE

[1]*Jimi Y.C. Wen,* [1]*Nikolay D. Gaubitch,* [2]*Emanuël A.P. Habets,* [3]*Tony Myatt and* [1]*Patrick A. Naylor*

[1]yung-chuan.wen@imperial.ac.uk

[1]Department of Electrical and Electronic Engineering, Imperial College London, SW7 2AZ, UK

[2]Department of Electrical Engineering, Technische Universiteit Eindhoven, P.O. Box 513, 5600 MB Eindhoven, The Netherlands

[3]Department of Music, University of York, Heslington, Y010 5DD, UK

ABSTRACT

Dereverberation is a growing area of research with many new algorithms appearing in the literature. However, there are still no unanimously accepted tools for evaluation of these algorithms. In this paper, we introduce the Multichannel Acoustic Reverberation Database at York (MARDY) containing real measured multichannel room impulse responses. We demonstrate its use for the evaluation of dereverberation algorithms using three recent multichannel methods. Furthermore, psychoacoustic issues regarding the performance evaluation of dereverberation algorithms are discussed.

1. INTRODUCTION

Reverberation occurs inside enclosed spaces, such as office rooms, due to the multipath propagation of acoustic signals from source to microphone. Reverberant speech can be described as sounding distant with noticeable colouration and echo. Whilst the effect is negligible with traditional handsets, reverberation affects the quality and intelligibility of speech in hands-free systems and is a significant problem for telecommunications, speech recognition applications and hearing aids [1]. Several dereverberation algorithms have been proposed and can be considered in two categories: (i) speech enhancement algorithms and (ii) blind channel estimation/inversion algorithms [1]. Dereverberation aim to form an estimate of the original source signal from one or more observed signals only. Multichannel processing is preferable, because it enables the use of spatial processing and provides more information about the source compared to single microphone approaches.

Objective and subjective measures are needed to evaluate the performance of dereverberation algorithms. In most of the current literature researchers use existing evaluation metrics such as SNR-based measures [2, 3] and spectral distortion based measures [3, 4, 5], which have been inherited from the speech enhancement and speech coding communities and for which the correlation with the perceived reverberation is not always clear. Alternatively, measures derived from the estimated room impulse response have been employed [6, 7] where the psychoacoustic effects are well understood. However, many speech dereverberation methods do not give a processed impulse response due to non-linear processing and thus making this type of evaluation difficult to use.The lack of unanimously accepted evaluation metrics for dereverberation makes results from different methods difficult to compare. Furthermore, it is desirable to use real measured room impulse responses in addition to rooms simulated using, for example the image method [8]. Even here some standardization would be useful, such that experiments can be repeatable and comparable.

In this paper, we introduce the Multichannel Acoustic Reverberation Database at York (MARDY) for evaluation of speech dereverberation algorithms. The database comprises a collection of room impulse responses measured with an eight element linear array for various source-array separations and for two different wall reflectivity settings. We also demonstrate the use of the database for three example algorithms using both subjective tests and objective measures.

The remainder of this paper is organized as follows. In Section 2 we describe the MARDY database. In Section 3 the three example algorithms are reviewed followed by various objective and subjective measures. Results from the objective measures and listening tests are discussed in Section 4, and finally conclusions are drawn in Section 5.

2. THE MARDY DATABASE

We now summarize the contents of and the measurement methodology for the construction of the MARDY database. The recordings were made in the Trevor Jones recording facility at the Music Research Centre of York University, UK. The recording room is a varechoic room with dimensions as specified in Fig. 1. The room was developed by Arup Acoustics and is controlled by a series of moving panels that can vary the acoustic properties of the

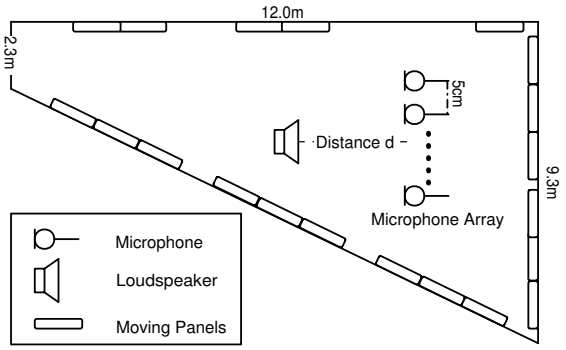

Figure 1:
Diagram of the recording room dimensions and setup. (Room Height=3 m)

room by changing the reflectivity of the walls. The facility has been constructed on suspended floors and has excellent acoustic isolation. A Genelec 1029A bi-amplified monitor system was used as source in the recording room. Two types of microphones were employed: (i) a single Schoeps Collette series microphone was used for reference and (ii) a microphone array comprising eight AKG C417 professional miniature condenser microphones. The separation between adjacent microphones was 5 cm and both loudspeaker and microphones were elevated to 1m above the floor.

We considered eight different acoustic systems consisting of four source-array separations and two moving panel configurations. Both Maximum Length Sequences (MLS) and speech were collected for each scenario. The MLS method [9, 10] was used to extract the impulse responses of the acoustic systems. The noise floor of the microphone array recordings was estimated to be -48 dB. The impulse responses can be used to obtain reverberant signals by convolution with anechoic recordings. A linear fade out was applied to the end of the estimated impulse responses in order to reduce the measurement noise in the tail. An example of a measured impulse response for one microphone at a distance of 4m from the source and for all reflective panels are shown in Fig. 2a. Figure 2b shows another example of a measured impulse response at a distance of 1m and for absorbent panels. The decay curves for the two settings are shown in Fig. 3a for the all reflective panel configuration and Fig. 3b for the all absorbent panel configuration. From these curves, it can be seen that the configuration of the moving panels dominates the shape of the decay curves, while increasing the distance of microphone to speaker decreases the clarity index (Table 1). In the case where the clarity index is large the reverberation tail effect will be masked by the stronger direct path component. We also observe that decay curves resulting from the same setup are all within 1 dB. We performed the following objective measures based on the impulse response: reverberation time $T60$, clarity index

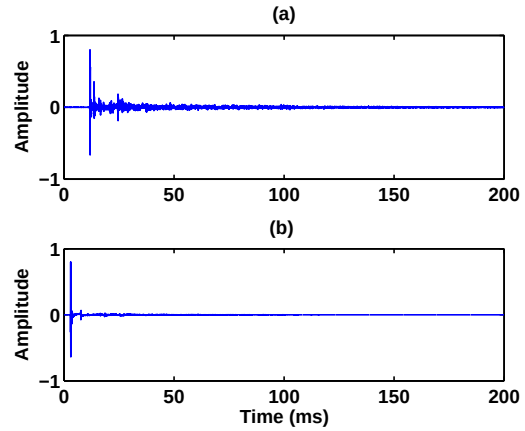

Figure 2:
Example of an estimated impulse response (a) for one microphone at a distance of 4m from the source with moving panels all reflective (b) for one microphone at a distance of 1m from the source with moving panels all absorbent.

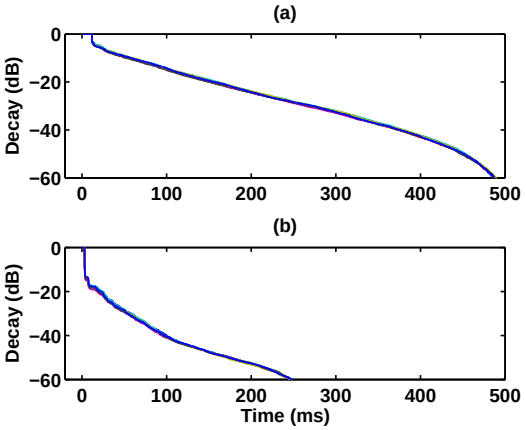

Figure 3:
Decay curves of the microphones array for (a) all reflective moving panels (4m) and (b) all absorbent (1m).

$C50$, the Deutlichkeit $D50$, and Centre Time TS [11]. The results averaged over all eight microphones are shown in Table 1. All the collected and processed data can be found at http://www.commsp.ee.ic.ac.uk/sap/.

3. EVALUATION OF SPEECH DEREVERBERATION ALGORITHMS USING MARDY

In this section, we demonstrate the application of the MARDY database in objective and subjective evaluation of speech dereverberation algorithms. We also examine some existing objective measures.

3.1. Dereverberation algorithms

We have selected three speech dereverberation methods for our example including: (i) a delay and sum beam-

Table 1:
Averaged objective measures of 8 acoustic systems

Distance	Panels	$T60$ ms	$C50$ dB	$D50$ %	TS ms
1 m	reflective	447	20.1	0.990	1.8
2 m	reflective	–	14.4	0.965	5.9
3 m	reflective	–	11.6	0.935	10.7
4 m	reflective	–	9.9	0.901	14.1
1 m	absorbent	291	29.4	0.999	0.7
2 m	absorbent	–	23.3	0.995	1.9
3 m	absorbent	–	20.7	0.992	3.0
4 m	absorbent	–	19.3	0.988	4.1

former (DSB)[12]; (ii) a multichannel approach based on a statistical model of late reverberation and spectral subtraction [5]; and (iii) a multi-microphone method using spatio-temporal averaging operating on the linear prediction residual [13]. We assume the positions of the source and microphones to be known for all three methods.

3.2. Objective Measures

We are interested in objective measures of dereverberation that only require the reverberant speech and the processed, dereverberated speech since the impulse response of the dereverberating system may not be obtainable. The following objective measures are examined using the corresponding anechoic speech [14] as reference: Segmental-SNR, Bark Spectral Distortion (BSD) [15], Cepstral distance (CD) [16], and Reverberation Decay Tail (R_{DT}) [17]. In our results we show the improvement in each measure indicated by the use of Δ.

3.3. Subjective Evaluation

Propagation of sound through an acoustic space causes two distinct perceptual effects [5, 17]: colouration, which results from frequency distortion due to the stronger early reflections, and reverberation decay tail effect, which results in temporal smearing due to the tail of the room impulse response. Colouration causes a sound to be boxy, thin, bright, etc., while the reverberation decay tail effect results in a distant and echo-ey sound quality. The subjective listening test was performed according to the guidelines of ITU-T Recommendation Series-P for subjective testing [18, 19]. Using the listening tests, we independently estimate the subjective perception of colouration (Col), reverberation decay tail effect (RTE), and the overall speech quality (Ovl). A total of 32 normal hearing subjects were recruited of which 26 gave results based on 64 speech files, with a male and a female talker for all eight acoustic systems, and the speech processed with the three dereverberation algorithms. Listeners indicated their rating on a 1-5 scale, with 5 being the best and 1 the worst for a given category. Calibration speech examples were given to assist listeners in identifying colouration and reverberation decay tail effect.

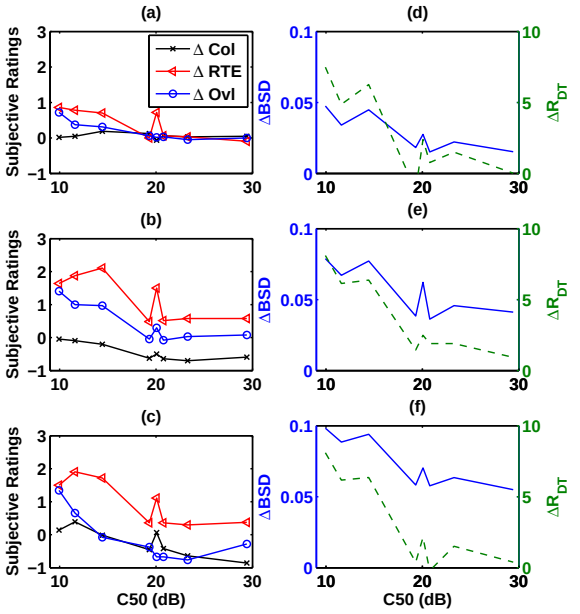

Figure 4:
Subjective measured performance for (a) DSB, (b) statistical model, (c) spatio-temporal averaging and objective performance masured with BSD (solid-line) and R_{DT} (dashed-line) for (d) DSB, (e) statistical model, (f) spatio-temporal averaging.

4. RESULTS

The results from the subjective (Col, RTE, Ovl) and the objective (BSD, R_{DT}) measurements are shown in Fig. 4 for a)d) DSB, b)e) statistical model and c)f) spatio-temporal averaging. It can be seen that all three algorithms reduce the effects of reverberation. As expected, the delay and sum beamformer only provides a limited amount of dereverberation, however, it also introduces the least processing distortions.

Table 2 shows the correlation for both the absolute and the difference of the subjective/objective measures. We can see that standard measures such as SNR, BSD and CD correlate poorly with perception of the three subjective qualities studied when evaluating different algorithms. The problem of these metrics is that they globally measure all types of effects reverberation, colouration, noise, etc. Thus these measures will be dominated by the results of the algorithms' different characteristics and other forms of distortion instead of measuring a distinct perceptual effect independently. The measure R_{DT} proposed in [17] shows higher correlation for RTE (0.62), and Δ R_{DT} shows higher correlation for ΔRTE (0.77) and ΔOvl (0.76). None of the measures investigated in this paper appears to have a good correlation for colouration.

Table 2:
Correlation between objective and subjective measures across the whole set of speech files (Reverberant and dereverberated)

	Δ Col	Δ RTE	Δ Ovl	Col	RTE	Ovl
ΔSNRseq	0.03	0.07	0.10	–	–	–
ΔBSD	0.15	0.51	0.15	–	–	–
ΔCD	0.13	0.51	0.24	–	–	–
$\Delta\ R_{DT}$	0.4	**0.77**	**0.76**	–	–	–
SNRseq	–	–	–	0.02	0.29	0.18
BSD	–	–	–	0.33	0.02	0.24
CD	–	–	–	0.40	0.14	0.17
R_{DT}	–	–	–	0.08	**0.62**	0.37

We additionally studied the relationship between the three subjective measures. The correlation between RTE and Col is 0.09, indicating that effects of reverberation tail and colouration can be considered substantially independent in our tests. There is higher correlation between the perception of RTE against Ovl (0.74) than Col against Ovl (0.38), showing the effect measured by RTE has greater importance in determining the overall speech quality in our application of interests.

5. CONCLUSION

We have presented a new database of measured room impulse responses for an eight element microphone array for four speaker-microphone configurations and different reverberation times. The use of the database for evaluation of dereverberation algorithms was illustrated with three example methods and for both subjective and objective measures. We investigated the perceptual effects of early and late reflections separately. The results from our listening tests show that many of the existing objective quality metrics do not perform well in predicting the perceptual effects of coloration. On the other hand the reverberation decay tail effect was captured by most of theses measures and in particular with R_{DT} showing good correlation and consistent results across the different algorithms.

6. REFERENCES

[1] P. A. Naylor and N.D. Gaubitch, "Speech dereverberation," in *Proc. Int. Workshop Acoust. Echo Noise Control*, 2005.

[2] M.J. Daly and J.R. Reilly, "Blind deconvolution using bayesian methods with application to the dereverberation of speech," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, 2004, vol. 2, pp. 1009–1012.

[3] J. Gonzalez-Rodriguez, J.L. Sanchez-Bote, and J. Ortega-Garcia, "Speech dereverberation and noise reduction with a combined microphone array approach," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, 2000, vol. 2, pp. 1037–1040.

[4] C. Marro, Y. Mahieux, and K. U. Simmer, "Analysis of noise reduction and dereverberation techniques based on microphone arrays with postfiltering," *IEEE Trans. Speech and Audio Process.*, vol. 6, no. 3, pp. 240–259, May 1998.

[5] E.A.P. Habets, "Multi-channel speech dereverberation based on a statistical model of late reverberation," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, 2005, vol. 4, pp. 173–176.

[6] T. Nakatani and M. Miyoshi, "Blind dereverberation of single channel speech signal based on harmonic structure," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, 2003, vol. 1, pp. 92–95.

[7] S. Weiss, G.W. Rice, and R.W. Stewart, "Multichannel equalization in subbands," in *Proc. IEEE Workshop on Applications of Signal Processing to Audio and Acoust.*, 1999, vol. 2, pp. 941–944.

[8] J.B. Allen and D.A. Berkley, "Image method for efficiently simulating small room acoustics," *J. Acoust. Soc. Amer.*, vol. 65, no. 4, pp. 943–950, 1979.

[9] M.R. Schroeder, "A new method of measuring reverberation time," *J. Acoust. Soc. Amer.*, vol. 37, pp. 409–412, 1965.

[10] J. Vanderkooy, "Aspects of MLS measuring systems," in *J. Audio Eng. Soc.*, 1994, vol. 42, pp. 219–231.

[11] H. Kuttruff, *Room Acoustics*, Taylor & Francis, 4 edition, Oct. 2000.

[12] B.D. Van Veen and K.M. Buckley, "Beamforming: a versatile approach to spatial filtering," *IEEE ASSP Magazine*, vol. 5, pp. 4–24, Apr. 1988.

[13] N.D. Gaubitch, P. A. Naylor, and D.B. Ward, "Multi-microphone speech dereverberation using spatio-temporal averaging," in *Proc. European Signal Process. Conference*, 2004, pp. 809–812.

[14] G. Lindsey, A. Breen, and S. Nevard, "SPAR's archivable actual-word databases," Tech. Rep., University College London, June 1987.

[15] S. Wang, A. Sekey, and A. Gersho, "An objective measure for predicting subjective quality of speech coders," *IEEE Journal on Selected Areas in Communications,*, vol. 10, no. 5, pp. 819 – 829, 1992.

[16] A.H. Gray Jr. and J.D. Markel, "Distance measures for speech processing," *IEEE Trans. on Acoustic, Speech and Signal Process.*, vol. 24, no. 5, pp. 381–390, Oct. 1976.

[17] J.Y.C. Wen and P.A. Naylor, "An evaluation measure for reverberant speech using tail decay modeling," *to appear in Proc. European Signal Process. Conference*, 2006.

[18] ITU-T, "Methods for subjective determination of transmission quality," in *Series P: Telephone Transmission Quality Recommendation P.800, ITU*, 1996.

[19] ITU-T, "Subjective test methodology for evaluating speech communication systems that include noise suppression algorithm," in *Series P: Telephone Transmission Quality Recommendation P.835, ITU*, 2003.