ROOMPRINTS FOR FORENSIC AUDIO APPLICATIONS

Alastair H. Moore, Mike Brookes and Patrick A. Naylor

Centre for Law Enforcement Audio Research
Department of Electrical and Electronic Engineering, Imperial College London, UK

ABSTRACT

A roomprint is a quantifiable description of an acoustic environment which can be measured under controlled conditions and estimated from a monophonic recording made in that space. We here identify the properties required of a roomprint in forensic audio applications and review the observable characteristics of a room that, when extracted from recordings, could form the basis of a roomprint. Frequency-dependent reverberation time is investigated as a promising characteristic and used in a room identification experiment giving correct identification in 96% of trials.

Index Terms— roomprint, acoustic impulse response, speech, forensic audio

1. INTRODUCTION

The field of forensic audio continues to grow in importance. Researchers in this field have hitherto concentrated on analysing speech recordings to identify the speaker (i.e. the "who") or the time it was made (i.e. the "when"). We now consider the task of identifying the acoustic environment (i.e. the "where").

It is well known that spaces sound different and effort has been dedicated to understanding those properties of a soundfield which lead to the differences in human perception (acoustic parameters), and the architectural structures which cause them (geometric features). Of course, a recording of speech will frequently pick up background noise from interfering sound sources (environmental sounds) which may be peculiar to the specific location. Thus, there is a plethora of information for forensic analysis.

To formalise this analysis we propose the concept of a *roomprint*. In an analogy to fingerprints, we envisage collecting a database of reference roomprints, equivalent to rolled fingerprints, against which a similar description derived from a recording, equivalent to a latent fingerprint, can be compared. This comparison will target two questions in particular:

1. Verification: If it is claimed that a recording was made in a particular room, is there sufficient evidence to reject the claim?
2. Identification: If it is known that a recording was made in one of a number of rooms, can we determine which one is most likely?

Question 2 has recently been considered for cataloging archive material [1]. Using MFCC-based classification of signals convolved with measured room impulse responses (RIRs), an equal error rate (EER) of 15% was obtained, indicating that these features capture some of the differences between the rooms. An analysis of the confusion matrix using multidimensional scaling showed that the first two dimensions of dissimilarity between the rooms were correlated with the early decay time and the bass ratio. However it was also observed that the rooms formed clusters according to the database they originated from, which suggests differences in measurement systems or technique may have caused some of the between-room variability.

For our intended application in law enforcement we would like to explicitly select the features of the roomprint based on quantifiable characteristics of the room. Hence, the contribution of this paper is three-fold. First we introduce the concept of a roomprint as a quantifiable description of an acoustic environment. Second we formulate a roomprint based on frequency-dependent reverberation time. Third we compare formulations based on alternative data transformations and show that taking the logarithm of reverberation time offers the best identification performance.

The remainder of this paper is organised as follows. In Sec. 2 we state the desired properties of a roomprint and consider the merits of geometric, acoustic and environmental information. In Sec. 3 we develop third-octave reverberation time as the basis for roomprints, present a room identification experiment and discuss the results. Finally, we draw conclusions in Sec. 4.

2. ROOMPRINTS

2.1. Requirements

A roomprint must exploit features of a room which allow it to be distinguished from other potentially similar rooms. A *reference roomprint* is obtained under ideal conditions with access to the room of interest. Thus it can include any aspect of the room which can be explicitly measured. On the other hand, a *latent roomprint* must be derived from an uncontrolled, usually single channel, recording of speech. Thus in classification, some measured features of a reference roomprint need to be inferred for a latent roomprint. The accuracy of this inference will dictate the utility of any employed feature.

A roomprint should ideally be invariant to the location of the talker and microphone in the room. Given that some variation is practically inevitable and the locations of talker and microphone are not necessarily known, the extent of variation expected should be quantified in the reference roomprint.

A roomprint should ideally be invariant with time. This may be harder to achieve than initially expected since many aspects of rooms are not time-invariant, such as the arrangement of furniture and the states of doors and windows. Moreover, the time of day will affect the types and level of environmental sound that are present in a recording.

2.2. Geometric features

The size and shape of a room are ideal features for inclusion in a roomprint because they are time-invariant and unrelated to the

talker or microphone position. They are also difficult features to infer from a single channel recording. Geometry inference has received a lot of attention recently but current methods typically require multiple microphones [2]. There is however some evidence to suggest that useful information about the volume of the space can be extracted from single channel impulse responses [3] and from reverberant speech [4]. Also, by making assumptions about the shape of the room, individual dimensions of a room can be inferred from a limited number of reflection time of arrivals [5]. Thus, system identification or reflection time of arrival estimation would enable the latent roomprint to include at least partial geometric features.

2.3. Room acoustics parameters

Room acoustics parameters are determined by the room geometry and the surface materials so are promising features . Within-room and between-room variations for five relatively uncorrelated parameters were reported in [6]. Early decay time (EDT) and reverberation time (T_{60}) are the most promising as they do not require a special microphone arrangement nor do they depend on the source directivity and orientation. The within-room standard deviation for T_{60} was found to be approximately half that of EDT in absolute terms. The extraction of T_{60} from reverberant speech is described in [7, 8]. However, the overall reverberation time alone is unlikely to be sufficient to distinguish between rooms. We therefore investigate frequency-dependent reverberation time. This is supported by the observation in [1] that the bass ratio (ratio of low and high frequency reverberation times) was a source of dissimilarity in their confusion matrix.

2.4. Environmental sounds

Environmental sounds (e.g. fans, building services) are not directly related to the properties of the room itself and often vary over time. Nevertheless, their presence in or absence from a speech recording could be used to verify a roomprint, especially if the date and time of the recording are known.

3. T_{60}-BASED ROOM IDENTIFICATION

To demonstrate the concept of roomprints, we next formulate the problem and illustrate roomprinting using a frequency dependent measure of reverberation time under a number of alternative transformations.

3.1. Problem formulation

The room impulse response (RIR) for room i measured at source-receiver configuration j is given by $h_{i,j}(t)$. T_{60} is the time taken for the acoustic energy in a room to decay by 60 dB after the source is turned off and it can be estimated from the slope of the normalised energy decay curve, which is found by reverse integrating the squared RIR [9].

According to Sabine's equation [10], T_{60} varies directly with the volume of the room, V, and inversely with the surface area of the room, S, and average absorption coefficient, α. Whilst V and S are determined by the geometry, α is determined by the surface material properties and is a function of frequency so it is common to calculate T_{60} separately within octave or $1/3$-octave bands. We define $\psi_{i,j,k}$ as the reverberation time in the $1/3$-octave frequency band k of room i, measured at source-receiver configuration j.

A single RIR measurement yields the reverberation time in each of K bands

$$\boldsymbol{\psi}_{i,j} = [\psi_{i,j,1}, \ldots, \psi_{i,j,K}]^T . \quad (1)$$

Combining the RIR measurements for J configurations of source-microphone position gives a $K \times J$ matrix

$$\boldsymbol{\psi}_i = \left[\boldsymbol{\psi}_{i,1}, \ldots, \boldsymbol{\psi}_{i,J}\right] \quad (2)$$

and combining measurements across I rooms gives an $K \times IJ$ matrix

$$\boldsymbol{\psi} = [\boldsymbol{\psi}_1, \ldots, \boldsymbol{\psi}_I] . \quad (3)$$

From a statistical point of view, each row of $\boldsymbol{\psi}_i$ represents a variable and each column an observation. Thus we can represent the observed values for a room using a K-dimensional multi-variate Gaussian distribution with mean $\boldsymbol{\mu}_i = [\mu_{i,1}, ..., \mu_{i,K}]^T$ and diagonal covariance matrix $\boldsymbol{\Sigma}_i = \text{diag}\left(\sigma^2_{i,1}, ..., \sigma^2_{i,K}\right)$.

Assuming that the reverberation times in different frequency bands are uncorrelated, the probability that a vector of $1/3$-octave band reverberation times $\boldsymbol{\psi}_{i',j'}$ observed at an unknown location j' in an unknown room i' was observed in a room with mean $\boldsymbol{\mu}_i$ and covariance $\boldsymbol{\Sigma}_i$ is given by

$$p(\boldsymbol{\psi}_{i',j'}|\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i) = \prod_{k=1}^{K} p(\psi_{i',j'}(k)|\boldsymbol{\mu}_i(k), \sigma^2_{i,k}) \quad (4)$$

$$= \prod_{k=1}^{K} \frac{1}{\sigma_{i,k}\sqrt{2\pi}} e^{\left(-\frac{\psi_{i',j'}(k)-\boldsymbol{\mu}_i(k)}{2\sigma^2_{i,k}}\right)} . \quad (5)$$

For the closed-set room identification task considered here, the room, i, which maximises $p(\boldsymbol{\psi}_{i',j'}|\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i)$ is chosen.

In (5) it is assumed that the variables used for room identification are the $1/3$-octave band reverberation times. However, these are not uncorrelated and may not follow a normal distribution. Therefore, we also consider transforming the data to some other representation, $\boldsymbol{\psi}'$, before fitting the statistical model. Six transformations are considered.

$$\boldsymbol{\psi}' = \boldsymbol{\psi} \quad (6)$$
$$\boldsymbol{\psi}' = \ln(\boldsymbol{\psi}) \quad (7)$$
$$\boldsymbol{\psi}' = \mathcal{KLT}(\boldsymbol{\psi}) \quad (8)$$
$$\boldsymbol{\psi}' = \mathcal{KLT}(\ln(\boldsymbol{\psi})) \quad (9)$$
$$\boldsymbol{\psi}' = \mathcal{DCT}(\boldsymbol{\psi}) \quad (10)$$
$$\boldsymbol{\psi}' = \mathcal{DCT}(\ln(\boldsymbol{\psi})) \quad (11)$$

where $\mathcal{KLT}(\cdot)$ denotes the Karhunen–Loève (KL) transform and $\mathcal{DCT}(\cdot)$ denotes the Discrete Fourier Transform (DCT).

Equation (6) is a direct mapping while (7) potentially gives a more even distribution by spreading out low values and compressing high ones.

The values of $\boldsymbol{\psi}_{i,j}$ are correlated (due to the smoothly varying nature of surface material absorption) so the assumption of independence in (5) is not valid. The KL transform used in (8) and (9) expresses a matrix as the coefficients of a set of basis-functions which are calculated such that the transformed coefficients are uncorrelated. As a result the variance in the data is compressed into the lower dimensions of the transformed coefficients.

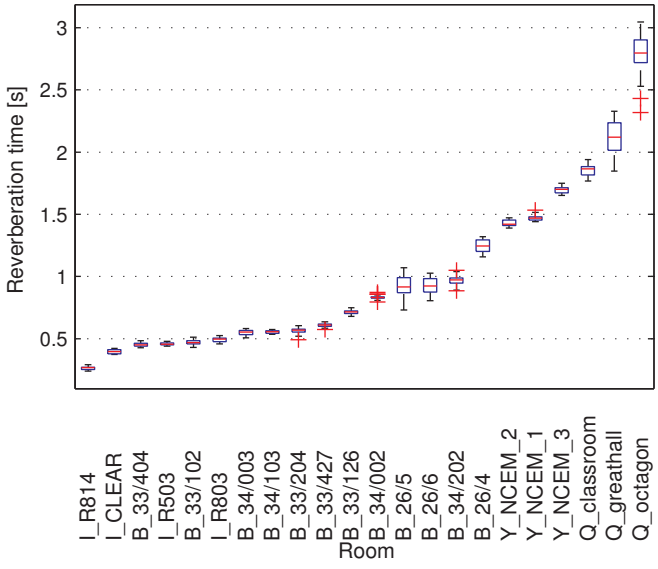

Figure 1: Distribution of T_{60} for each room in ascending order of mean T_{60}.

The KL transform is data-dependent such that the transformed coefficients used in a roomprint depend on which other roomprints are in the database. The DCT used in (10) and (11) expresses a matrix as the coefficients of data-independent orthogonal basis functions. This avoids the problem with the KLT but is sub-optimal in removing correlation between the variables.

3.2. Experiment

The RIRs for a total of 22 rooms, with volumes varying from 29 $m^3$to 9500 m^3, were sourced from a combination of publicly available data and measurements made by the authors. The distribution of T_{60} for each room is shown in Fig. 1. Room labels are prefixed to indicate their attribution: **B**GU [11], **Q**MUL [12], Uo**Y** [13]. Rooms prefixed **IC** were measured specifically for this study.

For each room 22 RIRs were selected, downsampled, where necessary, to 8 kHz and filtered into 14 $^1/_3$-octave bands with centre frequencies in the range 160 Hz to 3.15 kHz. From these $\boldsymbol{\psi}_i$ was estimated using the method from [14], which is more robust to noisy measurements than [9].

The dataset of 484 observations were transformed according to each of (6)-(11) to give six alternative representations. For each representation, a 14-dimensional Gaussian distributions was estimated for each room. A leave-one-out cross-validation procedure was used where, on each iteration, a single observation was omitted from the dataset used to train the models and assigned a (predicted) room label according to the maximum likelihood criterion. This was repeated for each observation in turn. A confusion matrix was compiled by counting the number of observations from each room (out of 22) that were assigned each of the possible labels. As a baseline comparison, the same cross-validation classification procedure was also applied to the overall reverberation time, T_{60}, as a one-dimensional feature.

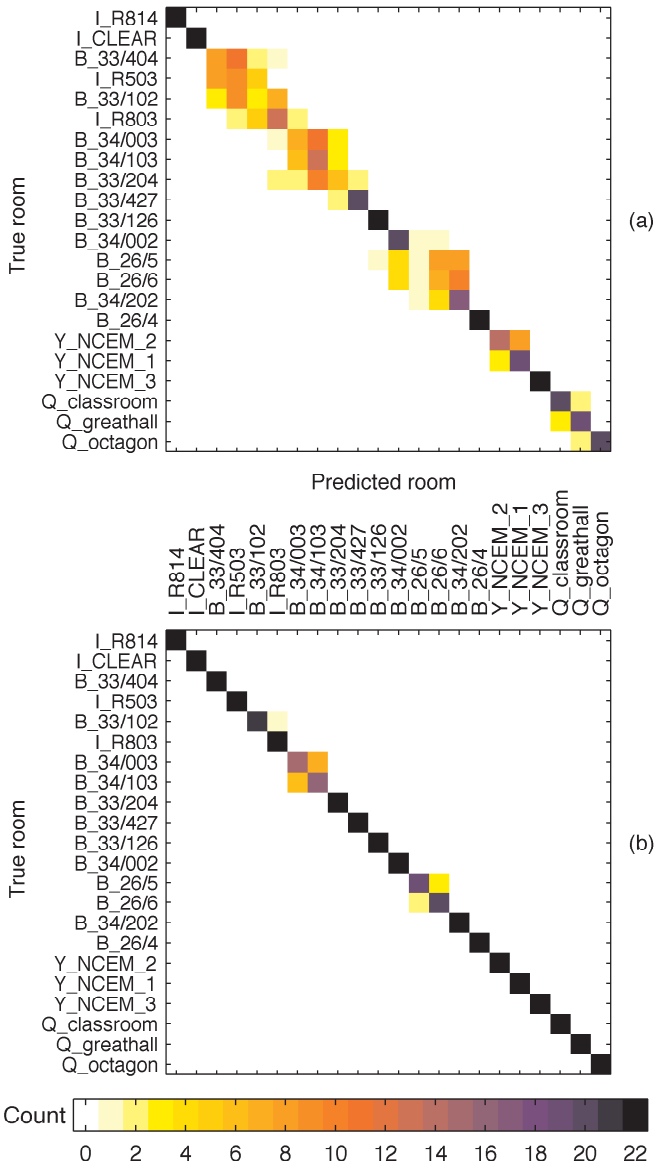

Figure 2: Confusion matrix for room identification experiment using (a) T_{60} and (b) $\ln(\psi)$ as features.

3.3. Results and Discussion

The confusion matrix for room identification based solely on T_{60} is shown in Fig. 2(a). The overall error rate is 32.6%. Most of the confusions are between rooms with similar mean T_{60}, with errors arising from mis-classification of rooms with similar T_{60} and the highest score in each room lies on the diagonal in all but 6 cases.

Using frequency-dependent reverberation time-based features the error rates were $\ln(\boldsymbol{\psi})$: 3.9%, $\boldsymbol{\psi}$: 4.1%, $\mathcal{KLT}(\boldsymbol{\psi})$: 5.4%, $\mathcal{KLT}(\ln(\boldsymbol{\psi}))$: 5.5%, $\mathcal{DCT}(\ln(\boldsymbol{\psi}))$: 6.6% and $\mathcal{DCT}(\boldsymbol{\psi})$: 7.6%. Fig. 2(b) shows the confusion matrix using the best of these, $\ln(\boldsymbol{\psi})$. The remaining identification errors are mostly caused by confusions between two pairs of rooms B_34/003 with B_34/103 and B_26/5 with B_26/6.

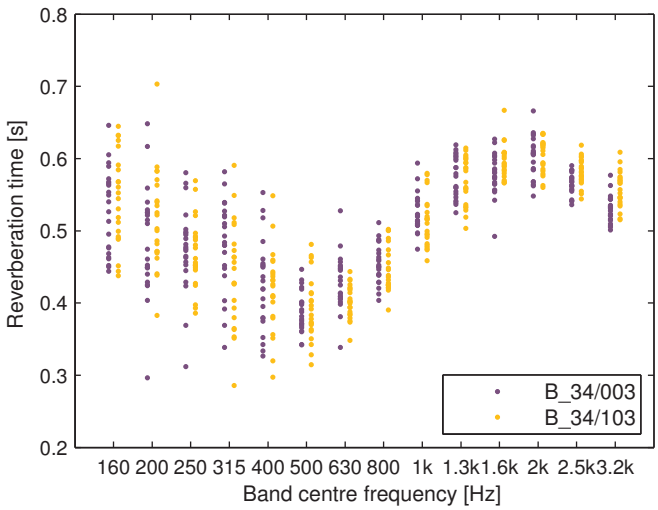

Figure 3: Comparison of distribution of ψ_i for rooms B_34/003 and B_34/103. Confusion in identification occurs in 30% of trials because they are very similar.

The measured reverberation times, ψ_i, for the confusable rooms B_34/003 and B_34/103 are shown in Fig. 3. We see that, since the two rooms are in the same building and are built to the same plan, they have almost identical distributions in each frequency bin. Despite this, our classifier is able to distinguish even these rooms correctly 70% of the time.

The fact that transforming the ψ data using either the KLT or the DCT was detrimental to classification performance is thought to be due to the fact that the variance in ψ_i is lower at higher frequencies than at low frequencies. Transforming into alternative domains, especially using the DCT, spreads the low frequency variance into all the dimensions of the transformed domain and so makes the distributions less separable.

4. CONCLUSIONS

We have presented the concept of roomprints in which a set of features of a room are inferred from a recording made in the room and are compared to a set of reference roomprints in order to perform identification or verification of the recording location. The frequency dependent reverberation time has been selected as a feature to illustrate the concept of roomprinting and a number of potential transformations of this feature set have been investigated. We have presented roomprinting results using the logarithm of frequency-dependent reverberation time as a roomprinting feature, for which an error rate of 3.9% has been obtained in a room identification experiment over 22 rooms.

5. REFERENCES

[1] N. Peters, H. Lei, and G. Friedland, "Name that room: Room identification using acoustic features in a recording," in *Proceedings of the 20th ACM international conference on Multimedia*, 2012, pp. 841–844.

[2] F. Antonacci, J. Filos, M. Thomas, E. Habets, A. Sarti, P. Naylor, and S. Tubaro, "Inference of room geometry from acoustic impulse responses," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 20, no. 10, pp. 2683 –2695, Dec. 2012.

[3] N. R. Shabtai, Y. Zigel, and B. Rafaely, "Feature selection for room volume identification from room impulse response," in *Proc. IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, 2009, pp. 249–252.

[4] ——, "Room volume identification from reverberant speech," in *Proc. Intl. Workshop Acoust. Echo Noise Control (IWAENC)*, 2010.

[5] A. H. Moore, M. Brookes, and P. A. Naylor, "Room geometry estimation from a single channel acoustic impulse response," in *Proc. European Signal Processing Conf. (EUSIPCO)*, Marrakech, Morocco, Sept. 2013, (submitted).

[6] X. Pelorson, J.-P. Vian, and J.-D. Polack, "On the variability of room acoustical parameters: Reproducibility and statistical validity," *Applied Acoustics*, vol. 37, pp. 175–198, 1992.

[7] N. D. Gaubitch, H. W. Löllmann, M. Jeub, T. H. Falk, P. A. Naylor, P. Vary, and M. Brookes, "Performance comparison of algorithms for blind reverberation time estimation from speech," in *Proc. Intl. Workshop Acoust. Signal Enhancement (IWAENC)*, Aachen, Germany, Sept. 2012.

[8] J. Eaton, N. D. Gaubitch, and P. A. Naylor, "Noise-robust reverberation time estimation using spectral decay distributions with reduced computational cost," in *Proc. IEEE Intl. Conf. on Acoustics, Speech and Signal Processing (ICASSP)*. Vancouver, Canada: IEEE, May 2013.

[9] ISO, *Acoustics-Measurement of the Reverberation Time of Rooms with Reference to Other Acoustical Parameters*, International Organization for Standardization (ISO) Recommendation ISO-3382, May 2009.

[10] H. Kuttruff, *Room Acoustics*, 4th ed. London: Taylor & Francis, 2000.

[11] N. R. Shabtai, Y. Zigel, and B. Rafaely, "Room volume identification from room impulse response using statistical pattern recognition and feature selection," vol. 128, pp. 1155–1162, 2010.

[12] R. Stewart and M. Sandler, "Database of omnidirectional and b-format room impulse responses," in *Proc. IEEE Intl. Conf. on Acoustics, Speech and Signal Processing (ICASSP)*, 2010, pp. 165–168.

[13] A. Foteinou, D. T. Murphy, and A. Masinton, "Verification of geometric acoustics-based auralization using room acoustics measurement techniques," in *Proc. AES Convention*, May 2010.

[14] M. Karjalainen, P. Antsalo, A. Mäkivirta, T. Peltonen, and V. Välimäki, "Estimation of modal decay parameters from noisy response measurements," *Journal Audio Eng. Soc.*, vol. 11, pp. 867–878, 2002.