19th INTERNATIONAL CONGRESS ON ACOUSTICS
MADRID, 2-7 SEPTEMBER 2007

INFORMATION TRANSFER IN AUDITORIA

PACS: 43.55.Hy

Summers, Jason E.
Naval Research Laboratory; Washington, D.C. 20375-5350, USA; jason.summers@nrl.navy.mil

ABSTRACT

It is hypothesized that subjective preference in auditoria is correlated with information transfer rate. Auditoria are considered as multiple-input multiple-output (MIMO) communication channels in which the spatial and directional distribution of the source ensemble and the spatial-hearing capacity of listeners are the mechanisms through which multipath increases information-transfer rate by overcoming finite spatial resolution. This framework makes predictions that are in agreement with prior phenomenological findings in the literature. In order to illustrate these predictions, channel capacity is calculated for computationally simulated auditoria.

INTRODUCTION

In Shannon's taxonomy [1], auditoria are acoustic channels for continuous communication systems of the type illustrated in Fig. 1.

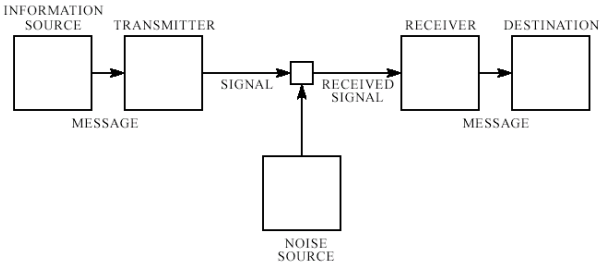

Figure 1.-Schematic representation of a continuous communication systems (from [1])

Signals $x(r'_i,t)$ are sent from one of n_T transmitters (acoustic sources) at r'_i to n_R=2 receivers (the ears of a listener) at r_j via a (possibly degenerate) multiple-input multiple-output (MIMO) waveform channel comprising acoustic propagation through the room, including the interaction of acoustic waves with the body (torso, head, and pinna) of the listener. The received signals $y(r_j,t)$ are expressed in terms of the room impulse responses (RIR) $h(r_j,r'_i,t)$ and noise signals at each receiver $w(r_j,t)$

$$y_j(t) = w_j(t) + \sum_i h_{ji}(t) * x_i(t), \qquad \text{(Eq. 1)}$$

or, in the frequency domain,

$$Y_j(f) = W_j(f) + \sum_i H_{ji}(f) X_i(f). \qquad \text{(Eq. 2)}$$

If subjective preference in auditoria has strong positive correlation with information transfer rate, the underlying goal of acoustical design of auditoria is to maximize the mutual information between the signal and received signal. Similarly, the suitability of a particular auditorium for a particular performance, whether speech or music, reflects the ability of the particular source-diversity scheme and space-time code represented by the performance style to optimize information transfer rate for the acoustic channel. While this viewpoint does not provide a

comprehensive framework–e.g., it may be aesthetically desirable to mask information in certain instances–it does provide insight into the interplay between room acoustics, signal characteristics, and subjective preference.

Information in auditoria

Following Wolfe [2], it possible to define three types of information encoded in the signal. The first is textual information, which consists of the categorically perceived, quantized variables as annotated in the score being performed or the text being read. The second is performance information, which consists of unannotated cues that allows distinction between different performers playing the same piece or prosodic cues that distinguish different orators reading the same text. The third is carrier information, which consists of additional information contributed by the source in the course of encoding the information for transmission. This includes not only inherent instrumental or vocal timbre but also information about location and directionality of source(s).

Explicit details of the encoding and decoding processes are not fully known, particularly the encoding and decoding of unannotated information. Both speech and music share common signal attributes, though allocate them differently between annotated and unannotated variables [2]. For speech, preference is largely (though not entirely) related to intelligibility, while, for music, preference is much less related to transcribability.

INFORMATION TRANSFER OVER SINGLE-INPUT SINGLE-OUTPUT CHANNELS

Many significant channel effects can be understood by considering the single-input single-output (SISO) channel between a single source-receiver pair. Shannon [1] distinguishes between two aspects of the channel that corrupt the signal: distortion and noise, the former being deterministic and the latter being stochastic. Distortion can, in principle, be removed by an inverse filter, while noise imposes a fundamental limit on channel capacity.

Noise is present in the acoustic channel in a variety of forms: first, as quasi-random background noise and, second, as aspects of the transmission path that are not or cannot be used to assist decoding the message. Convolution of the signal with the RIR of a reverberant room results in significant time spread at the receiver, which can be characterized by the time constant τ of the decaying energy envelope ε of the RIR

$$\varepsilon(t) = \varepsilon_0 \exp(-t/\tau), \qquad \text{(Eq. 3)}$$

where τ is the RMS delay spread of the channel, which is in general a function of frequency. Because convolution distortion is deterministic, it does not limit, in principle, the capacity of the channel. Rather, the capacity *C* (in bits/s) is governed only by the bandwidth *B* and the signal-to-noise ratio (SNR) at each frequency ([3], pp. 383-390),

$$C = \int_0^B \log_2\left(1 + \frac{|H(f)|^2}{W(f)}\right) df. \qquad \text{(Eq. 4)}$$

While, for a simple system consisting of the direct sound followed by a single reflection, Eq. (4) indicates a decrease in capacity as delay time and amplitude of the echo increase [4], it does not impose a fundamental limit due to the time spread of the RIR.

In contrast, it has been widely observed that excessive time spread resulting from large τ can compromise information transfer due to significant intersymbol interference (ISI). This apparent conflict is resolved by fundamental limitations on the ability of a receiver to deconvolve the RIR from the received signal.

First, reverberant rooms transition from deterministic behaviour in the early field to stochastic behaviour in the late field, the time-domain analogue to the transition that occurs in a transfer function above the Schroeder frequency [5]. Taking into account the limited time resolution of

the auditory system and the asymptotically quadratic growth of reflection density in time, it is possible to define a critical time

$$t_c = \sqrt{V}, \qquad \text{(Eq. 5)}$$

where V is room volume in cubic meters and t_c is critical time in milliseconds after which there are a sufficient number of reflections in each time-resolution cell that the impulse response can be regarded as a realization of a non-stationary stochastic process consisting of stationary coloured noise modulated by a frequency-dependent exponential envelope [5].

Second, real auditoria are characterized by time-variant multipath reflections that result in frequency-selective small-scale fading. Time variance arises through both fluctuations in the environment (e.g., air temperature) and small changes in the location or directional orientation of the source(s) and receivers. Such perturbations can result in relatively large changes in RIR through the near chaotic dependence of the RIR on boundary conditions [6]. Because this chaotic dependence is manifested primarily through accumulated uncertainty in subsequent reflections, low-order reflections in the early field are less affected by such perturbations than high-order reflections in the late field. For propagation delays longer than the coherence time of the channel (the time over which the RIR is static), channel estimation becomes intractable.

These physical limitations bound the channel-state information (CSI) available to the receiver and are consistent with observations that listeners are insensitive to details of late-reverberation fine structure [7]. This implies that listeners cannot fully estimate and deconvolve the RIR from the received signal, due to excessive time spread and/or time variance, and therefore the rate of information transfer is limited by uncompensated ISI. Thus, the late portion of the RIR is effectively noise while the early portion either is not detrimental or even beneficial, as will be shown.

Capacity and intelligibility

In the presence of noise, increased reverberation brings a trade-off between increased SNR and increased ISI [8, 9]. For speech, ISI results from overlap-masking due to temporal spread of one phoneme into the following phoneme [9]. These observations lead to the concept of an optimal reverberation time that depends not only on room volume, source power, noise-source power, and source-receiver locations [8], but also, as Bolt [9] has shown, the statistical distribution of phonemes in time. As confirmation, it is observed that reducing speaking rate mitigates most of the adverse effects of reverberation on intelligibility [9]. All of this is consistent with the interpretation of Eq. (4) presented above: increased intelligibility is realized by increasing SNR without resulting in uncorrected ISI that reduces the effective SNR. The often-observed effect of slowing down of speech in a reverberant room is an adaptation that attempts to maximize the amount of information retrievable from the received signal.

INFORMATION TRANSFER OVER MULTIPLE-INPUT MULTIPLE-OUTPUT CHANNELS

Multipath gain through spatial diversity

MIMO channels provide the possibility of both array gain and diversity gain [10]. If the sources have no CSI, each source is allocated equal power and the MIMO system realises only diversity gain via space-time coding [11]. Assuming that the receivers have perfect CSI the spectral efficiency of the channel in bits/s/Hz is

$$C = \log_2 \det\left(\mathbf{I}_{n_R} + \frac{\mathrm{SNR}}{n_T}\mathbf{H}\mathbf{H}^*\right) = \sum_{k=1}^{\max\{n_T, n_R\}} \log_2\left(1 + \frac{\mathrm{SNR}}{n_T}\lambda_k^2\right), \qquad \text{(Eq. 6)}$$

where λ_k are the singular values of $\mathbf{H}$. Equation (6) indicates that the number of receivers must be greater than or equal to the number of sources to decode all of the eigenmodes of a full-rank channel matrix. In a multipath channel, the rank of the channel matrix is related to the number of image sources [12].

For the concert-hall problem, Eq. (6) is disadvantageous because it allocates power between the sources as though there were n_T eigenchannels even though there are only n_R [10]. Multipath diversity can yield even greater capacity gains if the transmitters have full or partial

CSI. If the channel is known, the power is distributed among the sources by water-filling in order to take full advantage of array gain for each of the subchannels in addition to the diversity gain. As Fig. 2 shows, this yields some advantage for $n_T > n_R$ [10].

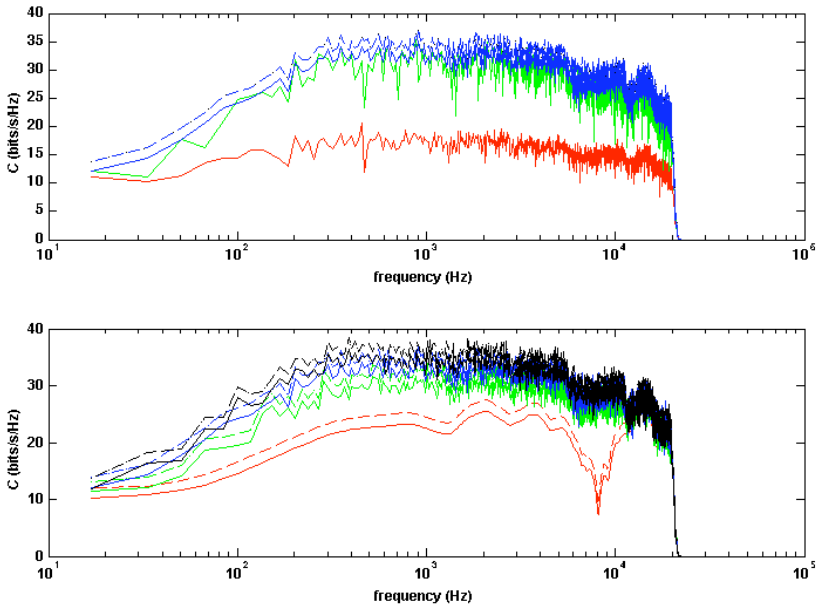

Figure 2.-Spectral efficiency calculated from computed RIR given uniform power (solid) and water-filling (dotted) for increasing n_T in a room with $13.82\tau = 2$ s and SNR = 60 dB (upper figure) for $n_T = 1$ (red), $n_T = 2$ (green) and $n_T = 4$ (blue); and given fixed $n_T = 4$, $n_R = 2$ with SNR=60 dB (lower figure) for increasing $13.82\tau = 0$ s (red), 1 s (green), 2 s (blue), and 3 s (black).

In auditoria, perfect CSI is impossible for both the source(s) and receivers. However, both have estimates of statistical properties such as τ and the receivers have knowledge of the early deterministic portion of the RIR. Partial CSI is optimal if it directs transmission along those eigenchannels that are known [13]. While Eq. (6) assumes that the receivers have perfect CSI, it has been shown that this is not required for slow fading (a random channel that is fixed over the symbol period) [14].

Increased spatial information through multipath

Prior work on localization in rooms has focused on those cases of small source-receiver separations for which source locations are well resolved under anechoic conditions and reverberation (both early and late reflections) degrades the accuracy of spatial judgments by distortion of spectral and temporal cues (see, e.g., [15]). In contrast, the cases considered here involve large source-receiver distances (small subtended angles) for which scattering from the head alone is insufficient to fully resolve source locations.

Following Buck [16] and Gaumond [17], the spatial information encoded by the acoustic scattering described in a set of impulse responses can be quantified. Given a set of n_T binaural room impulse responses (BRIR) formed by concatenating the RIR measured at the left and right ears, they can be expressed in terms of a bilinear expansion [3]

$$h_m(t) = \sum_{n=1}^{N} \alpha_{mn} \varphi_n(t), \qquad \text{(Eq. 7)}$$

where the orthonormal functions $\varphi_n(t)$ are the right singular vectors of the SVD. Additive white Gaussian noise representing perceptual noise can be similarly expanded in the same $\{\varphi_n(t)\}$ with zero-mean iid coefficients.

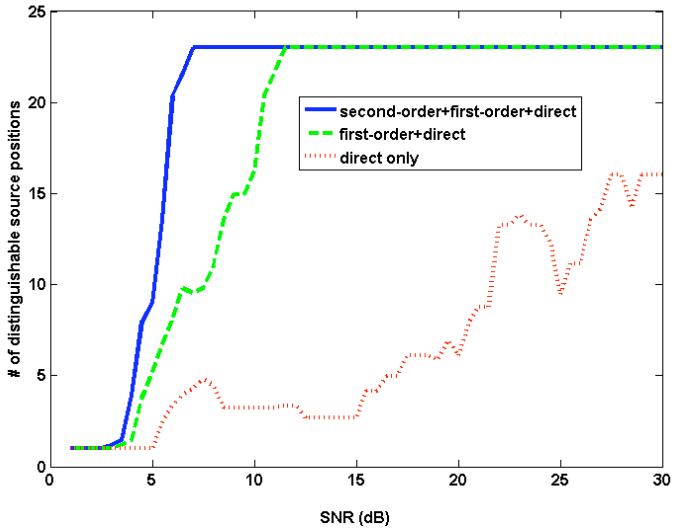

Figure 3.-The number of distinguishable source positions as a function of SNR for 23 sources subtending angle of 58 deg is calculated from computed BRIR for increasing reflection orders.

The coefficients are quantized with a discretization interval ρ that is a multiple of the noise variance

$$\underline{c}_m = \left[\mathrm{int}\left(\frac{\alpha_{m1}}{\rho}\right) \cdots \mathrm{int}\left(\frac{\alpha_{mN}}{\rho}\right) \right]. \qquad (\text{Eq. 8})$$

The spatial information contained in the coefficients is given by the sum of the probability of each coefficient vector

$$E_{out} = \sum_{\underline{c}} P(\underline{c}) \log_2 P(\underline{c}), \qquad (\text{Eq. 9})$$

where

$$P(\underline{c}) = \sum_{m|\underline{c}=\underline{c}}^{N} p_m, \qquad (\text{Eq. 10})$$

expressed in terms of the probability of each input position that results in that coefficient vector. If each vector of discrete coefficients is unique to a single input position, the input information is equal to the output information. However, degeneracy of the coefficient vectors, due either to noise or the field, reduces the output information and results in spatially unresolved sources.

Scattering is the key to overcoming the finite angular resolution of the receiver and lifting this degeneracy [11, 12], as shown in Fig. 3. While it is not clear that listeners can exploit this additional information, there is evidence that listeners can learn echo patterns in rooms, minimizing the degradation of spatial judgments of fully resolved sources [18].

DISCUSSION AND CONCLUSIONS

The model presented here is in accord with the notion that every performance is suited to a particular τ and is (implicitly) composed with this τ in mind ([19] and [20], pp. 21-23 and 535-536). Likewise, it affirms that the preferred reverberation time of an auditorium should be covariant with the size of the ensemble ([20], p. 22).

Spatially, the model suggests that larger values of lateral fraction are preferable, particularly in the early field, because such values indicate an increase in the angular spread of the incident sound field and therefore the number of encodable channels and the information capacity of the

system. Similarly, the model suggests that low values of inter-aural cross correlation will be preferred because less correlation between the signals arriving at the two receivers maximizes the capacity of the channel.

Lubman [21] has proposed that changes in liturgical styles that evolved along with church architecture reflected the need to provide optimum encoding for the message. Similarly, the history of Western music can be cast as the macro co-evolution of musical styles and auditorium architecture. Arrangement of the orchestra by a conductor can be viewed as microevolution that configures the channel to better suit the space-time code of a particular piece.

Anecdotally, auralizations of auditoria are often perceived as more reverberant than the real room they are modelling despite accurate calculated reverberation time. The model presented here suggests that this may be partly due to unrealistic modelling of sources (e.g., as a single point rather than spatially distributed), which fails to exploit multipath diversity.

ACKNOWLEDGEMENTS
This work was supported by the Office of Naval Research. Portions were conducted while J.E.S was a visiting researcher with the Advanced Systems Development Centre, Yamaha Corporation of Japan.

References:
[1] C. E. Shannon: A Mathematical Theory of Communication. Bell Systems Technical Journal **27**, (1948) 379-423 and 623-656; corrected version obtained from http://cm.bell-labs.com/cm/ms/what/shannonday/paper.html
[2] J. Wolfe: Speech and music, acoustics and coding, and what music might be 'for'. Proceedings of the 7th International Conference on Music Perception and Cognition, Sydney (2002) 10-13
[3] R. G. Gallager: Information Theory and Reliable Communication. (Wiley, New York, 1968) 355-441
[4] D. R. Hummels: The Capacity of a model for the underwater acoustic channel. IEEE Transactions on Sonics and Ultrasonics **SU-19, No.3** (1972) 350-353
[5] J.-D. Polack: La transmission de l'énergie sonore des les salles. Thèse de Doctorat d'Etat, Université du Maine, Le Mans
[6] F. Fahy: Foundations of Engineering Acoustics. (Academic Press, New York, 2001) 273
[7] K. Meesawat and D. Hammershoi: An investigation on the transition from early reflections to a reverberation tail in a BRIR. Proceedings of the International Conference on Auditory Display, Kyoto (2002) 1-5
[8] R. H. Bolt and A. D. MacDonald: Theory of speech masking by reverberation. Journal of the Acoustical Society of America **21**, (1949) 577-580
[9] M. Hodgson and E.-M. Nosal: Effect of noise and occupancy on optimal reverberation times for speech intelligibility in classrooms. Journal of the Acoustical Society of America **111**, (2002) 931-939
[10] J. B. Andersen: Array gain and capacity for known random channels with multiple element arrays at both ends. IEEE Journal on Selected Areas in Communications **18, No. 11** (2000) 2172-2178
[11] G. J. Foschini and M. T. Gans: On limits of wireless communication in a fading environment when using multiple antennas. Wireless Personal Communications **6, No.3** (1998) 311-335
[12] P. F. Driessen and G. J. Foschini: On the capacity formula for multiple input-multiple output wireless channels: a geometric interpretation. IEEE Transactions on Communications **47, No.2** (1999) 173-176
[13] E. Jorswieck and H. Boche: Optimal transmission with imperfect channel state information at the transmit antenna array. Wireless Personal Communications **27** (2003) 33-56
[14] T. L. Marzetta and B. M. Hochwald: Capacity of a Mobile Multiple-Antenna Communication Link in Rayleigh Flat Fading. IEEE Transactions on Information Theory **45, No.1** (1999) 139-157.
[15] B. G. Shinn-Cunningham, N. Kopco, and T. J. Martin: Localizing nearby sound sources in a classroom: Binaural room impulse responses. Journal of the Acoustical Society of America **117**, (2005) 3100-3115
[16] J. R. Buck: Information theoretic bounds on source localization performance. Proceedings of the IEEE Sensor Array and Multichannel Signal Processing Workshop, Washington, DC (2002) 184-188
[17] C. F. Gaumond: Broadband information transfer from oceanic sound transmission. Acoustic Research Letters Online **5, No.2** (2004) 44-49
[18] B. G. Shinn-Cunningham and N. Kopco: Effects of reverberation on spatial auditory performance and spatial auditory cues (A). Journal of the Acoustical Society of America **111**, (2002) 2440
[19] Y. Ando, T. Okano, and Y. Takazoe: The running autocorrelation function of different music signals relating to preferred temporal parameters of sound fields. Journal of the Acoustical Society of America **86**, (1989) 644-649
[20] L. L. Beranek: Concert and Opera Halls: How They Sound. (Acoustical Society of America, Woodbury, NY, 1996)
[21] D. Lubman and B. H. Kiser: The History of Western Civilization Told Through the Acoustics of its Worship Spaces. Proceedings of the 17th International Congress on Acoustics, Rome (2001)