On Perceptual Audio Compression with Side Information at the Decoder

Adel Zahedi*, Jan Østergaard*, Søren Holdt Jensen*, Patrick Naylor†, and Søren Bech*‡
*Department of Electronic Systems
Aalborg University, 9220 Aalborg, Denmark
Email: {adz, jo, shj, sbe}@es.aau.dk
†Electrical and Electronic Engineering Department
Imperial College, London SW7 2AZ, UK
Email: p.naylor@imperial.ac.uk
‡Bang & Olufsen
7600 Struer, Denmark

Abstract

Due to the distributed structure of many modern audio transmission setups, it is likely to have an observation at the receiver which is correlated with the desired source at the transmitter. This observation could be used as side information to reduce the transmission rate using distributed source coding. How to integrate distributed source coding into the perceptual audio compression procedure is thus a fundamental question. In this paper, we take a completely analytical approach to this problem, in particular to the rate-distortion trade-off and the corresponding coding schemes. We then interpret the results from an audio coding perspective. The main result is that, to upgrade a regular perceptual audio coder to a distributed coder, one needs to revise the perceptual masking curve. The revised masking curve models the availability of the side information as an extra masking effect, yielding lower rates. Interestingly, this means that at least conceptually, the distributed coding scenario could be integrated into the audio coder with minor changes, and without destructing the original coder.

The research leading to these results has received funding from the European Union's Seventh Framework Programme (FP7/2007-2013) under grant agreement n° ITN-GA-2012-316969.

I. Introduction

It is a well-known fact among the audio coding community that sound quality as assessed by human could not be adequately represented by the traditional fidelity criteria such as the mean-squared error. This was best demonstrated by an experiment performed at the Bell Labs in the late eighties. Two noisy versions of an audio signal with the same measured SNR but with differently shaped noises in the spectral domain were presented to the listeners. In one version, the noise was white, and in the other one, it was perceptually shaped. While the listeners found the noise in the first version annoying, the one in the second version was assessed as inaudible or barely noticeable [1].

The fact that the auditory system's ability to detect the noise in a given audio excerpt depends on the frequency and also the excerpt itself has lead to efficient audio compression strategies based on the perceptual shaping of the coding noise. This has been an active area of research for decades and has lead to popular and highly efficient audio coding standards such as the MPEG audio [2].

Due to the increasing interest in the distributed and networked scenarios for multimedia transmission, it is likely in modern audio coding setups, that there is already a slightly different version of the sound source at the decoder. Consider for example the case where a microphone records its observation of a sound source and transmits to a center where there is another microphone. Here due to the correlation between the recordings, the data from the second microphone could be used as the side information at the decoder. This allows to reduce the transmission rate without compromising the quality using Distributed Source Coding [3]. Here is thus the fundamental question to ask: How could one integrate the distributed source coding techniques into the perceptual audio coding procedure?

The simplest way is to encode the audio source exactly as the non-distributed scenario, and then use the availability of the side information for binning based on the Slepian and Wolf's results in [4] to reduce the transmission rate [5]. However, such an approach is merely based on intuition, and there is no argument to justify why one should apply the perceptual masking curve exactly the same way as the non-distributed coding scenario.

In this paper, we start with the fundamentals. For analytical derivations, we assume the Gaussian distribution for the sources. We derive the rate-distortion functions for the distributed coding scenario as the bounds on the best performance, and suggest an optimal coding scheme that could be used to achieve these bounds. By studying the implications of this coding scheme, we show that the availability of the side information at the decoder could be modelled as an extra masking effect. The distributed perceptual coding scenario thus requires the application of a revised masking curve which depends on the the side information as well as the perceptual masking curve. The important conclusion regarding the implementation is that, in principle, the only changes that have to be made in a regular perceptual audio coder to upgrade it to a distributed audio coder is to revise the perceptual masking curve before applying it, and then add a binning step to the output of the encoder.[1]

Although the Gaussian distribution may not be an accurate model for audio signals, it is a well-known fact that for most setups, it is the worst distribution for coding (see e.g.

[1] In some cases it may turn out that this step adds too much complexity for only slightly improving the performance. In this case, it may be skipped.

[6], [7]), and coding schemes derived based on the Gaussianity assumption lead to no worse results, if the assumption does not hold. Moreover, as a sanity check for the results derived in this paper based on the Gaussianity assumption, we consider the regular audio coder as a special case, and show that our results are in agreement with what actual coders do.

We denote random processes by boldface lower-case letters, e.g. $\mathbf{y}$, where for simplicity of notation, we have skipped the dependency on time. The power spectral density of $\mathbf{y}$ is denoted by $S_y(f)$. More generally, we use S with and appropriate subscript to denote the conditional and nonconditional spectral and cross-spectral densities. As an example, $S_{yx|uz}$ is the conditional cross-spectral density of $\mathbf{y}$ and $\mathbf{x}$ given $\mathbf{u}$ and $\mathbf{z}$. Markov chains are denoted by two-headed arrows, e.g. $\mathbf{u} \leftrightarrow \mathbf{y} \leftrightarrow \mathbf{z}$.

The rest of the paper is organized as follows. In Section II, we give a very brief introduction to perceptual audio coding. In Section III, the source coding problem is formulated and solved to derive the rate-distortion functions and achievable coding schemes. Section IV is dedicated to the interpretation of the results of Section III from an audio coding perspective. Section V concludes the paper.

II. Perceptual Audio Coding

Digital audio is the result of quantization of audio signals, and is inevitably corrupted by coding noise. For perceptually transparent audio coding, the noise added to the audio signal due to coding has to be inaudible.

The human auditory system is not equally sensitive to all frequencies. Figure 1 illustrates the sound pressure level (SPL) at the absolute threshold of hearing as a function of frequency. As seen from the figure, the minimum SPL for which the sound is audible is much higher at low and high frequencies compared to e.g. frequencies between 3 and 4 kHz. Coding noise with a flat spectrum and of a certain level may thus be inaudible at low or high frequencies, but audible and perceptually annoying at medium frequencies. However, this is not the whole story. As another well-known fact, as illustrated in Fig. 2, the availability of the sound at a certain frequency creates a masking effect around that frequency, such that the threshold of hearing increases locally. This means that around the frequency of a masker, the level of the coding noise could be higher without being audible. Moreover, since audio is highly nonstationary, the maskers change amplitude and location by time, implying that the shaping of the coding noise has to be frame-dependent.

Observations of this type with the root in Psychoacoustics gave rise to the theory of perceptual audio coding. A typical perceptual audio coder combines all the local masking effects due to the tonal and nontonal maskers at the current audio frame together with the absolute threshold of hearing to form the frame-dependent global perceptual masking curve as a function of frequency. If the audio frame is quantized such that the coding noise is below this masking curve at all frequencies, the noise will not be audible. A perceptual audio coder with a given rate thus allocates the available bit pool to different frequency components of the audio frame based on this maximum inaudible noise level. One could for example normalize the audio spectrum by the global masking curve in the frequency domain, and then quantize the result uniformly. The interested reader is referred to [1] for more details on perceptual audio coding principles and standards.

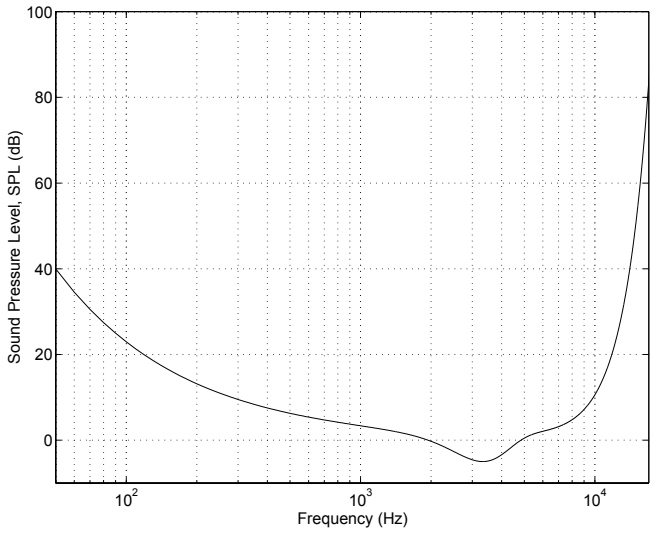

Figure 1: Absolute threshold of hearing in quiet for an average listener (from [8])

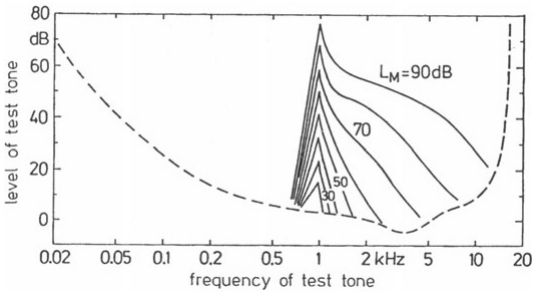

Figure 2: Level of test tones just masked by 1 kHz tones of different levels (from [9]). The dashed curve illustrates the absolute threshold of hearing. A tone of a level below the solid curve corresponding to an available masker will not be audible.

III. Source Coding Problem

The block diagram of the source coding problem is shown in Fig. 3. $\mathbf{y}$ and $\mathbf{z}$ are jointly Gaussian stationary random processes which are observed at the encoder and decoder, respectively. The encoder encodes the observation and sends the message $\mathbf{u}$ to the decoder. The decoder receives the message from the encoder and makes an estimation $\hat{\mathbf{y}}$ of the desired source $\mathbf{y}$ using the received message and the side information $\mathbf{z}$. The problem is to find the minimum rate R such that for a given target spectral distortion function $S_D(f)$, the spectrum $S_{y|uz}(f)$ of the reconstruction error at the decoder satisfies:

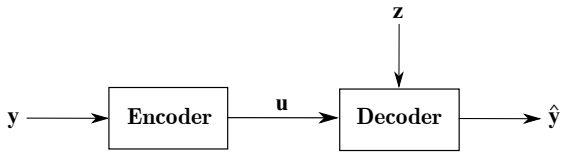

Figure 3: Block diagram of the coding system with side information at the decoder

$$S_{y|uz}(f) \leq S_D(f), \tag{1}$$

at all frequencies f. This problem can be formulated as the following minimization problem:

$$R\left(S_D(f)\right) = \min_{\mathbf{u}} I\left(\mathbf{y}; \mathbf{u}|\mathbf{z}\right) \quad \text{s.t.} \quad \begin{cases} S_{y|uz}(f) \leq S_D(f), \\ \mathbf{u} \leftrightarrow \mathbf{y} \leftrightarrow \mathbf{z}. \end{cases} \tag{2}$$

Theorem III.1. *The rate-distortion function $R\left(S_D(f)\right)$ defined in (2) is given by:*

$$R\left(S_D(f)\right) = \frac{1}{2} \int \log \frac{S_{y|z}(f)}{\min\left(S_{y|z}(f), S_D(f)\right)} df, \tag{3}$$

and is achieved by the following linear coding scheme:

$$\mathbf{u}^* = \mathbf{y} + \boldsymbol{\nu}, \tag{4}$$

where the coding noise $\boldsymbol{\nu}$ is Gaussian and uncorrelated with $\mathbf{y}$, and has the following power spectral density:

$$S_\nu(f) = \left(\frac{1}{\min\left(S_D(f), S_{y|z}(f)\right)} - \frac{1}{S_{y|z}(f)} \right)^{-1}. \tag{5}$$

Moreover, the spectrum of the reconstruction error at the optimal decoder is given by:

$$S_{y|u^*z}(f) = \min\left(S_D(f), S_{y|z}(f)\right). \tag{6}$$

Proof: See the appendix. ■

Note that (4) means that $\mathbf{y}$ should be quantized such that the dequantized version is given by (4); i.e., the resulting coding noise $\boldsymbol{\nu}$ is Gaussian and independent of $\mathbf{y}$ with the spectrum given by (5). This can be achieved by applying a dithered vector quantizer. Also notice that the rate-distortion function (3) is a generalized form of the the case with no side information in [10, Theorem 1].

Finally, rewriting (4) in the spectral domain yields:

$$S_{u^*}(f) = S_y(f) + S_\nu(f). \tag{7}$$

IV. Perceptual Coding Interpretations

The distortion constraint in (1) together with the proof of Theorem III.1 implies that the estimate $\hat{\mathbf{y}}' = E[\mathbf{y}|\mathbf{u}, \mathbf{z}]$ satisfies the following:

$$\mathbf{y} = \hat{\mathbf{y}}' + \mathbf{e}, \tag{8}$$

such that $\mathbf{e}$ is independent of $\hat{\mathbf{y}}'$, and

$$S_e(f) = \min\left(S_D(f), S_{y|z}(f)\right), \tag{9}$$

where (9) follows from (6). Writing the linear estimation of $\hat{\mathbf{y}}'$ in terms of $\mathbf{y}$ in (8) yields:

$$\hat{\mathbf{y}}' = \mathbf{h} * \mathbf{y} + \mathbf{e}', \tag{10}$$

where $*$ denotes the convolution operation, $\mathbf{e}'$ is independent of $\mathbf{y}$ with the following spectrum:

$$S_{e'}(f) = \frac{S_e(f)}{S_y(f)}\left(S_y(f) - S_e(f)\right), \tag{11}$$

and the Fourier transform $H(f)$ of $\mathbf{h}$ is given by:

$$H(f) = \frac{S_y(f) - S_e(f)}{S_y(f)}. \tag{12}$$

Deconvolving $\mathbf{y}$ from $\hat{\mathbf{y}}'$ in (10) and denoting the result by $\hat{\mathbf{y}}$, one could rewrite (10) as:

$$\hat{\mathbf{y}} = \mathbf{y} + \mathbf{n}, \tag{13}$$

where we have:

$$S_n(f) = \left(S_e^{-1}(f) - S_y^{-1}(f)\right)^{-1}. \tag{14}$$

Note that $\hat{\mathbf{y}}$ in (13) is the reproduced audio at the decoder and $\mathbf{n}$ is the added noise due to the compression/decompression process. In order for the noise to be inaudible at a given frequency f, we must have $S_n(f) \leq S_M(f)$, where $S_M(f)$ is the global masking curve at frequency f.

Let us assume that we would like to be on the borderline of transparency, and thus $S_n(f) = S_M(f)$. To achieve this particular choice of $S_n(f)$, one needs to specify a particular target spectral distortion that yields $S_n(f) = S_M(f)$. Let us denote this target distortion by $S_D^*(f)$. The minimum rate required to achieve transparency would then be $R(S_D^*(f))$. We substitute $S_D^*(f)$ in (9) to obtain the corresponding $S_e^*(f)$. From (14) it follows that:

$$S_e^*(f) = \min\left(S_D^*(f), S_{y|z}(f)\right) = \left(S_y^{-1}(f) + S_M^{-1}(f)\right)^{-1}, \tag{15}$$

To encode the audio stream for this target distortion (and thus to achieve the minimum rate for transparency), we use the achievable scheme in (7). We substitute (15) in (5), and the result in (7) to obtain:

$$S_{u^*}(f) = S_y(f) + \frac{S_M(f)}{1 - S_M(f)\left[S_{y|z}^{-1}(f) - S_y^{-1}(f)\right]} \tag{16}$$

$$= S_y(f) + S'_M(f). \tag{17}$$

where

$$S'_M(f) = \frac{S_M(f)}{1 - S_M(f)\left[S^{-1}_{y|z}(f) - S^{-1}_y(f)\right]}. \tag{18}$$

Equation (16) is the core of this work and is discussed more elaborately in the sequel.

A. No Side Information

Suppose that $\mathbf{z} = \mathbf{0}$, which means that there is no side information at the decoder, and the problem is simplified to a normal audio coding problem. From (18) it follows that in this case, we have $S'_M(f) = S_M(f)$, and thus the coding scheme in (16) is reduced to:

$$S_{u^*}(f) = S_y(f) + S_M(f). \tag{19}$$

Dividing the two sides of (19) by $S_M(f)$ yields:

$$\frac{S_{u^*}(f)}{S_M(f)} = \frac{S_y(f)}{S_M(f)} + 1. \tag{20}$$

Equation (20) implies that one could first normalize the spectrum of the audio signal with the masking curve.[2] The result should then be quantized uniformly, since the quantization noise in (20) is fixed. This is similar to what typical perceptual audio coders do, and thus could also be considered as a sanity check for the above results.

B. Distributed Coding Case

Here due to the presence of the side information at the decoder, the perceptual masking curve $S_M(f)$ should be replaced by an *equivalent mask* $S'_M(f)$ which is a revised version of $S_M(f)$. Similar to the previous case, one could normalize the spectrum of the signal by the equivalent masking curve $S'_M(f)$, and then uniformly quantize the result.

Note that from the fact that $S_{y|z}(f) \leq S_y(f)$ and (18), it follows that:

$$S'_M(f) \geq S_M(f), \tag{21}$$

which means that the availability of the side information is equivalent to an extra masking effect, which allows for higher quantization noise (yielding thus lower rates or higher compression ratios) without compromising the quality of the reconstructed audio. It is noteworthy though that this equivalent masking curve is not a perceptual phenomenon. The extra coding noise added due to the higher masking levels suggested by the equivalent

[2]In Sound Pressure Level domain, it means that one should subtract the masking curve from the signal spectrum.

masking curve would not be perceptually masked at the decoder and would be audible unless the decoder makes use of the side information to compensate for it.

The optimal estimation of the audio signal at the decoder using the received data and the side information should be performed based on the proof of Theorem III.1 (see in particular (26) in the appendix).

Finally, we would like to emphasize that perceptual audio coding is deeply involved in heuristics. Although it is supported by strong psychoacoustical arguments, there have been several issues regarding the implementation which have been partially resolved during the past few decades. It would be a very long way to go, if distributed perceptual audio coding would require starting from the scratch. It is therefore a significant advantage that in principle the implementation of a distributed coder based on the above results would only require an extra step where the perceptual masking curve is revised using (18).[3]

V. Conclusions

We studied the problem of distributed perceptual audio coding. We started with an information theoretic analysis of the problem. We formulated the coding problem as a rate-distortion problem with a power spectral density distortion constraint that models the perceptual masking curve for the audio coder. For this problem, we derived the rate-distortion function and optimal coding schemes from which we inferred the implications of the theoretical results from an audio coding perspective. Most particularly, we showed that the distributed coding scenario requires merely a revision of the perceptual masking curve to an equivalent masking curve that in addition to the perceptual masking effects, takes into account also the availability of the side information at the decoder. This paper though is merely a preliminary report of the concept. Applying the results to simple audio examples is the work in progress. Future work could be considering more complex audio sources, and eventually building up an actual distributed perceptual audio coder.

Appendix
proof of Theorem III.1

We first lowerbound the rate-distortion function defined by (2) with (3), and then upperbound it with the same function. The combination of the lower and upper bounds thus gives the exact rate-distortion function.

To lowerbound $R(S_D(f))$ in (2), we write the following chain of inequalities:

$$\begin{aligned} R(S_D(f)) &= \min_{\mathbf{u}} I(\mathbf{y};\mathbf{u}|\mathbf{z}) \quad \text{s.t.} \quad \begin{cases} S_{y|uz}(f) \le S_D(f), \\ \mathbf{u} \leftrightarrow \mathbf{y} \leftrightarrow \mathbf{z}. \end{cases} \\ &\ge \min_{\mathbf{u}} I(\mathbf{y};\mathbf{u}|\mathbf{z}) \quad \text{s.t.} \quad S_{y|uz}(f) \le S_D(f) \qquad (22) \\ &= \min_{\mathbf{u}} h(\mathbf{y}|\mathbf{z}) - h(\mathbf{y}|\mathbf{u},\mathbf{z}) \quad \text{s.t.} \quad S_{y|uz}(f) \le S_D(f) \\ &\ge \min_{S_{y|uz}(f)} \frac{1}{2}\int \log \frac{S_{y|z}(f)}{S_{y|uz}(f)} df \quad \text{s.t.} \quad \begin{cases} S_{y|uz}(f) \le S_D(f) \\ S_{y|uz}(f) \le S_{y|z}(f) \end{cases} \qquad (23) \end{aligned}$$

[3]Note that the final binning step should be applied to the output of the encoder and will not interfere with the coding procedure.

$$= \frac{1}{2}\int \log \frac{S_{y|z}(f)}{\min\left(S_D(f), S_{y|z}(f)\right)} df$$

where (22) is because removing a constraint enlarges the search space, and (23) is because the Gaussian distribution maximizes the differential entropy. Note that the additional constraint $S_{y|uz}(f) \leq S_{y|z}(f)$ in (23) is necessary, because any valid conditional power spectral density $S_{y|uz}(f)$ must satisfy this constraint.

To upperbound $R(S_D(f))$, we propose a particular choice of $\mathbf{u}$ denoted by $\mathbf{u}^*$ which satisfies the Markov chain in (2), i.e. $\mathbf{u} \leftrightarrow \mathbf{y} \leftrightarrow \mathbf{z}$, and has the following two properties:

1) The required rate for delivering $\mathbf{u}^*$ to the decoder is no more than $R(S_D(f))$.
2) The reconstruction error at the decoder given $\mathbf{u}^*$ and $\mathbf{z}$ has a spectrum $S_{y|u^*z}(f)$ which satisfies the distortion constraint $S_{y|u^*z}(f) \leq S_D(f)$.

We will show that $\mathbf{u}^*$ in (4) has these two properties. To prove the first one, based on the results in [4] and [11], it is enough to show that $I(\mathbf{y}; \mathbf{u}^*|\mathbf{z})$ gives (3). Noting that $\mathbf{u}^*$ in (4) is Gaussian, we write:

$$\begin{aligned} I(\mathbf{y}; \mathbf{u}^*|\mathbf{z}) &= h(\mathbf{u}^*|\mathbf{z}) - \mathbf{h}(\mathbf{u}^*|\mathbf{y}, \mathbf{z}) \\ &= h(\mathbf{y} + \boldsymbol{\nu}|\mathbf{z}) - h(\boldsymbol{\nu}) \\ &= \frac{1}{2}\int \log \frac{S_{y|z}(f) + S_{\nu}(f)}{S_{\nu}(f)} df \qquad (24) \\ &= \frac{1}{2}\int \log \frac{S_{y|z}(f)}{\min\left(S_D(f), S_{y|z}(f)\right)} df, \qquad (25) \end{aligned}$$

where (25) follows from substituting (5) in (24).

To prove the second property, without loss of generality, we write $\mathbf{y}$ in terms of its linear estimation from $\mathbf{u}^*$ and $\mathbf{z}$ as follows: [4]

$$\mathbf{y} = (\mathbf{h}_1 * \mathbf{u}^*) + (\mathbf{h}_2 * \mathbf{z}) + \mathbf{e}, \qquad (26)$$

where the estimation error $\mathbf{e}$ is independent of $\mathbf{u}^*$ and $\mathbf{z}$, with the spectrum $S_e(f) = S_{y|u^*z}(f)$, and based on the independence of $\mathbf{e}$ from $\mathbf{u}^*$ and $\mathbf{z}$, one can derive the following formulas for the Fourier transform of the linear filters $\mathbf{h}_1$ and $\mathbf{h}_2$:

$$\begin{aligned} H_1(f) &= \frac{S_z(f)S_y(f) - |S_{yz}(f)|^2}{S_z(f)S_u(f) - |S_{yz}(f)|^2} \\ H_2(f) &= \frac{S_{yz}(f)S_{\boldsymbol{\nu}}(f)}{S_z(f)S_u(f) - |S_{yz}(f)|^2}. \end{aligned}$$

Using (26), we have:

$$S_{y|z}(f) = S_{u^*|z}(f)\,|H_1(f)|^2 + S_e(f),$$

[4] From this it follows that the best estimation of $\mathbf{y}$ given $\mathbf{u}^*$ and $\mathbf{z}$ is $\hat{\mathbf{y}}' = (\mathbf{h}_1 * \mathbf{u}^*) + (\mathbf{h}_2 * \mathbf{z})$.

from which it follows that:

$$
\begin{aligned}
S_{y|u^*z}(f) = S_e(f) &= S_{y|z}(f) - S_{u^*|z}(f)\,|H_1(f)|^2 \\
&= S_{y|z}(f) - S_{y|z}(f)H_1^*(f) && (27)\\
&= S_{y|z}(f) - \frac{|S_{y|z}(f)|^2}{S_{u^*|z}(f)} && (28)\\
&= S_{y|z}(f) - \frac{S_{y|z}^2(f)}{S_{y|z}(f) + S_{\boldsymbol{\nu}}(f)} && (29)\\
&= S_{y|z}(f) - \frac{S_{y|z}^2(f)}{S_{y|z}(f) + \left(\frac{1}{\min\left(S_D(f), S_{y|z}(f)\right)} - \frac{1}{S_{y|z}(f)}\right)^{-1}} && (30)\\
&= \min\left(S_D(f), S_{y|z}(f)\right) \leq S_D(f),
\end{aligned}
$$

where (27) is because from (4) and (26) we have:

$$S_{y|z}(f) = S_{yu^*|z}(f) = S_{u^*|z}(f)H_1(f), \tag{31}$$

and (28), (29) and (30) follow respectively from (31), (4) and (5). The proof is now complete.

REFERENCES

[1] M. Bosi and R. E. Goldberg, *Introduction to digital audio coding and standards*, Kluwer Academic Publishers, Second printing, 2003.

[2] *The Moving Picture Experts Group*, http://mpeg.chiariglione.org/, accessed Nov. 2015.

[3] A. Zahedi, J. Østergaard, S. H. Jensen, P. Naylor, and S. Bech, *Distributed remote vector Gaussian source coding for wireless acoustic sensor networks*, IEEE Data Compression Conference, Snowbird, UT, Mar. 2014.

[4] D. Slepian and J.Wolf, *Noiseless coding of correlated information sources*, IEEE Transactions on Information Theory, vol. 19, no. 4, pp.4711-480, Jul. 1973.

[5] A. Majumdar, K. Ramchandran, and I. Kozintsev, *Distributed coding for wireless audio sensors,* IEEE Workshop on Applications of Signal Processing to Audio and Acoustics, New Paltz, NY, USA, Oct. 2003.

[6] T. Berger, *Rate-distortion theory,* Wiley Online Library, USA, 2003.

[7] A. Zahedi, J. Østergaard, S. H. Jensen, P. Naylor, and S. Bech, *Audio coding in wireless acoustic sensor networks*, Signal Processing, vol. 107, pp. 141–152, Feb. 2015.

[8] T. Painter and A. Spanias, *Perceptual coding of digital audio*, Proceedings of the IEEE, vol. 88, no. 4, pp. 451-515, 2000.

[9] H. Fastl and E. Zwicker, *Psychoacoustics, facts and models*, Springer, 3rd edition, 2007.

[10] Y. Kochman, J. Østergaard, and R. Zamir, *Noise-Shaped Predictive Coding for Multiple Descriptions of a Colored Gaussian Source*, IEEE Data Compression Conference, Snowbird, UT, pp. 362-371, Mar. 2008.

[11] A. Wyner and J. Ziv, *The rate-distortion function for source coding with side information at the decoder*, IEEE Transactions on Information Theory, vol. 22, no. 1, pp. 1-10, Jan. 1976.