Impact of Noise Correlation on Multimodality

Saloua Chlaily, Pierre-Olivier Amblard, Olivier Michel, Christian Jutten

HAL Id: hal-01382832
https://hal.archives-ouvertes.fr/hal-01382832

Impact of Noise Correlation on Multimodality

Saloua Chlaily, Pierre-Olivier Amblard, Olivier Michel and Christian Jutten

Gipsa-Lab, CNRS, Université Grenoble Alpes, France

Abstract—In this paper, we consider the problem of estimating an unknown random scalar observed by two modalities. We study two scenarios using mutual information and mean square error. In the first scenario, we consider that the noise correlation is known and examine its impact on the information content of two modalities. In the second scenario we quantify the information loss when the considered value of the noise correlation is wrong. It is shown that the noise correlation usually enhances the estimation accuracy and increases information. However, the performance declines if the noise correlation is misdefined, and the two modalities may jointly convey less information than one single modality.

I. INTRODUCTION

Recently, in various domains, different sensors *i.e.* modalities are used simultaneously to observe several aspects of the phenomenon studied, hopefully ensuring a better understanding: this concept is called Multimodality. Multimodality exploits the links between modalities, to enable their interaction and therefore reveal more information about the phenomenon than using a unique modality. However, the expected improvement is not always observed. That is because multimodality introduces new challenges: different resolutions, sizes, missing data, inconsistency, to name a few, and noise. Noise is one of the challenges rarely addressed since most studies ignore it or consider independent noises [1]. However, the correlation between noises, if it exists, represent an important link that must be considered for ensuring better performance.

The use of different modalities is important in many applications where phenomena are extremely complex (e.g. *neuroimaging, human-computer interface, remote sensing, etc*). Furthermore, the benefit of using different modalities is very application dependent. Important challenges in multimodal signal processing are not only to process jointly several modalities, but also to *a priori* quantify the benefits (if any) of multimodal integration. The work presented here is a small step in this direction. To draw conclusions independently of the applications, we consider a model which is simple to analyze, easy to interpret, but that yet provides interesting conclusions about the effect of interaction between modalities onto inference tasks.

The model involves only two modalities and a few parameters. Considering each modality as a communication channel, we address the following questions in terms of mutual information and of minimum mean square error (MMSE).

1) Under which conditions, a single modality carry more information than two modalities?

2) What is the importance of noise correlation on inference accuracy?
3) What is the effect of misdefining the links between noises on the information content of modalities and inference accuracy?

The model considered in this work, was studied in [2] in terms of Fisher information, while still under the exact model assumption. The paper is organized as follows. The proposed model and main notations are presented in Section II. In Section III, we recall some notions of information theory. The impact of the links between modalities on the information content is presented in Section IV. In Section V we investigate the effect of misdefining the links between modalities. Finally, the conclusions and perspectives are drawn in Section VI.

II. NOTATIONS AND MODEL

Let X_1 and X_2 be two noisy observations of a random scalar parameter S. For simplicity, we ignore the sensor response:

$$\begin{cases} X_1 = S + B_1 \\ X_2 = S + B_2 \end{cases}, \tag{1}$$

where S and the noises B_1 and B_2 are assumed to be jointly Gaussian, with zero means and variances σ^2, σ_1^2, and σ_2^2, respectively. S is independent of the noises. We denote by $\rho = \frac{E[B_1 B_2]}{\sigma_1 \sigma_2}$ the noise correlation coefficient, and by $R_i = \frac{\sigma^2}{\sigma_i^2}$ the signal to noise ratio (SNR) of modality X_i. We assume without loss of generality that $R_1 \geq R_2$, unless otherwise stated.

III. UNCERTAINTY OF EXACT AND WRONG MODELS

In this section, we introduce the differential entropy (DE) and the mutual information (MI) [3], which we will use in the next sections for studying the model (1).

The DE quantifies the uncertainty of a continuous random variable, *i.e.* the information that would be obtained by observing this variable. Let X be a continuous random variable with a probability density function (pdf) f_X, the DE of X denoted by $H(X)$ is given by:

$$H(X) = \int f_X(x) \log \frac{1}{f_X(x)} dx. \tag{2}$$

Eq. (2) assumes that the pdf f_X is known. In fact, if we do not know the exact probability of X, and assume a wrong pdf $\hat{f}_X$ instead of f_X, the measure of Shannon information is function of $\hat{f}_X$, while the average is over the exact probability:

$$H_{\hat{f}_X|f_X}(X) = \int f_X(x) \log \frac{1}{\hat{f}_X(x)} dx. \tag{3}$$

The notation $H_{\hat{f}_X|f_X}(X)$ denotes the expected uncertainty of the random variable X when we believe that its pdf is $\hat{f}_X$ while it is f_X. In coding theory, for discrete variables, $H_{\hat{f}_X|f_X}(X)$ is the number of bits needed to code an event when its probability density function is wrongly estimated. If we write $H_{\hat{f}_X|f_X}$ as a function of H, we find:

$$H_{\hat{f}_X|f_X}(X) = H(X) + D_{KL}\big(f_X(X)||\hat{f}_X(X)\big), \quad (4)$$

where $D_{KL}\big(f_X(X)||\hat{f}_X(X)\big) = \int f_X(x)\log\frac{f_X(x)}{\hat{f}_X(x)}dx$ is the Kullback-Leibler (KL) divergence between the densities f_X and $\hat{f}_X$. Since D_{KL} is always nonnegative and is zero iff $\hat{f}_X = f_X$ [3], (4) leads to the following interpretation: when we assume $\hat{f}_X \neq f_X$, the uncertainty of X increases and the error $H_{\hat{f}_X|f_X} - H$ is given by the KL divergence between exact and wrong densities. Clearly, the error on the uncertainty increases with the KL divergence i.e with the pdf mismatch.

Let Y and Z be two random variables linked to a third X. Observing Y and Z reduces the uncertainty of X. The amount of information that Y and Z reveal jointly about X is given by the MI:

$$\begin{aligned}I\big(X;(Y,Z)\big) &= H(X) - H(X \mid Y,Z) \\ &= D_{KL}\big(f_{XYZ}(x,y,z)||f_X(x)f_{YZ}(y,z)\big),\end{aligned}(5)$$

where $H(X \mid Y,Z) = -E_{f_{XYZ}}\log f_{X|YZ}(x \mid y,z)$ is the conditional DE that quantifies the remaining uncertainty of X given (Y,Z). $f_{X|YZ}(x \mid y,z)$ is the conditional pdf of the random vector X given (Y,Z). $f_{XYZ}(x,y,z)$ and $f_{YZ}(y,z)$ are the joint densities for (X,Y,Z) and (Y,Z) respectively. If we consider a wrong conditional pdf for (Y,Z) given X, $\hat{f}_{YZ|X}(y,z \mid x)$, the amount of information revealed jointly by Y and Z about X would be wrongly evaluated as:

$$\begin{aligned}I_{\hat{f}_{YZ|X}|f_{YZ|X}}&\big(X;(Y,Z)\big) \quad\quad\quad\quad\quad\quad\quad (6)\\ &= H(X) - H_{\hat{f}_{YZ|X}|f_{YZ|X}}(X \mid Y,Z)\\ &= I\big(X;(Y,Z)\big) - D_{KL}\big(f_{X|YZ}(x \mid y,z)||\hat{f}_{X|YZ}(x \mid y,z)\big)\end{aligned}$$

where:

$$\hat{f}_{X|YZ}(x \mid y,z) = \frac{\hat{f}_{YZ|X}(y,z \mid x)f_X(x)}{\hat{f}_{YZ}(y,z)}$$

and $\quad \hat{f}_{YZ}(y,z) = \int \hat{f}_{YZ|X}(y,z \mid x)f_X(x)dx$.

Clearly (6) shows that the MI decreases when we assume a wrong model, due to the property of D_{KL}, and the equality $I_{\hat{f}_{YZ|X}|f_{YZ|X}} = I$ holds iff $\hat{f}_{YZ|X} = f_{YZ|X}$. Note that $I_{\hat{f}_{YZ|X}|f_{YZ|X}}\big(X;(Y,Z)\big)$ is *not* a mutual information, in the sense that it has not the properties of MI. In fact, MI is a KL divergence (nonnegative), while $I_{\hat{f}_{YZ|X}|f_{YZ|X}}$ is a difference of KL divergences and it can be negative (cf. Section V).

IV. EXACT MODEL

In this section we assume that the link between noises is perfectly known, i.e the exact value of the noise correlation coefficient ρ is known. The modalities in (1) can be considered
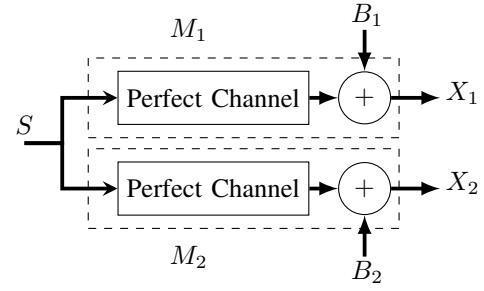

Fig. 1. Modalities as two communication channels

as two noisy communication channels (Fig. 1), where S is the input and X_1, X_2 are the outputs.

Using the Gaussian assumption (Section II), the information carried by the unique modality X_i ($i = 1,2$) about S is

$$I(S;X_i) = \frac{1}{2}\log(1 + R_i). \quad (7)$$

Following the assumption $R_1 \geq R_2$ *i.e.* X_1 is more reliable than X_2, X_1 provides more information about S, *i.e.* $I(S;X_1) \geq I(S;X_2)$. If the modalities are considered jointly, i.e by exploiting their link ρ, the information they provide about S is given by (5), and in the Gaussian case it is given by:

$$\begin{aligned}I\big(S\,;&(X_1,X_2)\big)\\ &= \frac{1}{2}\log\left(\frac{(1+R_1)(1+R_2) - (\sqrt{R_1R_2}+\rho)^2}{1-\rho^2}\right). \quad (8)\end{aligned}$$

Eq. (8) depends on the SNRs and on the parameter ρ which summarizes the relation between the two modalities. Thus, in the following we add ρ as a subscript to distinguish between the MI for different values of ρ. The notation $I_\rho\big(S;(X_1,X_2)\big)$ denotes the MI between S and the couple of modalities (X_1,X_2) when the noise correlation coefficient is ρ. In Figure 2, we draw $I_\rho\big(S;(X_1,X_2)\big)$ (plain), $I_0\big(S;(X_1,X_2)\big)$ (dashed) and $I(S,X_1)$ (dotted) as a functions of ρ for equal and different SNRs (Fig. 2a and 2b, respectively), while the interpretations are presented in the following subsections.

From an estimation theory perspective, the optimal MMSE estimator [4] of S as a function of X_1 and X_2 is:

$$\hat{S}_\rho(X_1,X_2) = \frac{(R_1 - \sqrt{R_1R_2}\rho)X_1 + (R_2 - \sqrt{R_1R_2}\rho)X_2}{(1+R_1)(1+R_2) - (\sqrt{R_1R_2}+\rho)^2}. \quad (9)$$

Fig. 3 shows the mean square errors (MSE) of estimations using only one modality, $E\Big[\big(S - \hat{S}(X_1)\big)^2\Big]$ (dotted), and the two modalities $E\Big[\big(S - \hat{S}_\rho(X_1,X_2)\big)^2\Big]$ (plain) vs. the noise correlation coefficient ρ. The interpretations are given in the following subsections while the additional dashed line, $E\Big[\big(S - \hat{S}_{0|\rho}(X_1,X_2)\big)^2\Big]$ is discussed in Section V.
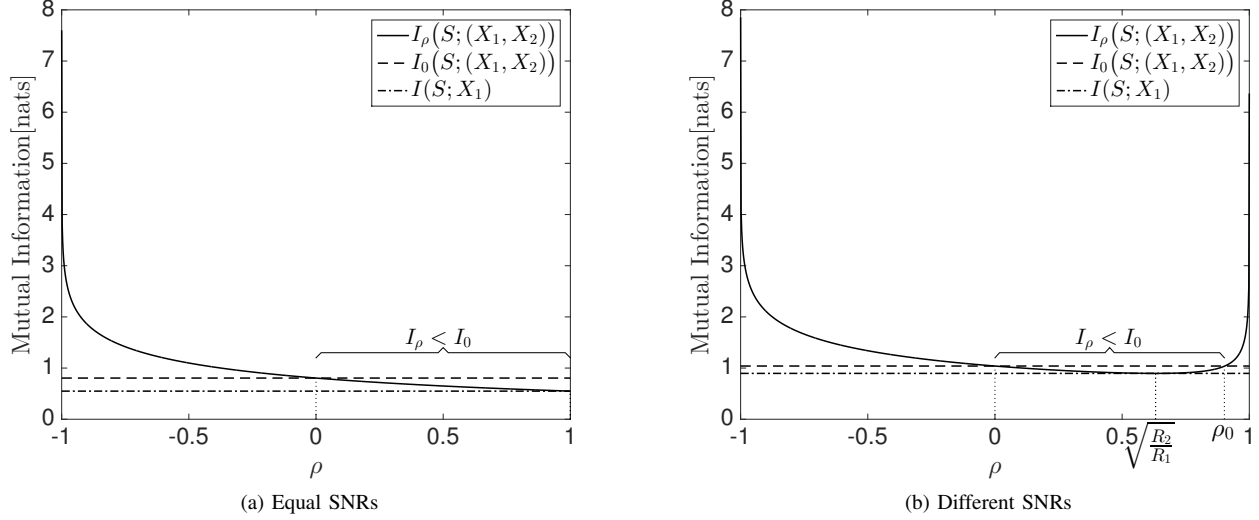
Fig. 2. The mutual information $I_\rho\big(S;(X_1,X_2)\big)$ (plain), $I_0\big(S;(X_1,X_2)\big)$ (dashed) and $I(S;X_1)$ (dotted) as a functions of ρ (known)

A. One or two modalities

First we would like to know when it is advantageous to combine two modalities. In this purpose, we compare the information brought jointly by the modalities with the information conveyed by a single modality:

$$I_\rho\big(S;(X_1,X_2)\big) - I(S;X_i) = I_\rho(S;X_j \mid X_i), \ i \neq j \quad (10)$$

where $I_\rho(S;X_j \mid X_i) = H(S \mid X_j) - H(S \mid X_i,X_j)$ is the conditional MI that quantifies the additional information carried by X_j given X_i. The conditional MI is always non-negative, and it equals zero when no further information can be retrieved from the new modality X_j.

For model (1), from equations (7) and (8), the conditional MI writes:

$$I_\rho(S\ ;X_i \mid X_j)$$
$$= \frac{1}{2} \log \left(\frac{(1+R_1)(1+R_2) - (\sqrt{R_1R_2} + \rho)^2}{(1-\rho^2)(1+R_j)} \right) \ (11)$$

$I_\rho(S;X_1 \mid X_2) \neq 0$ since X_1 provides more information than X_2 about S ($R_1 > R_2$). On the contrary, the conditional MI $I_\rho(S;X_2 \mid X_1)$ can vanish for $\rho = \sqrt{\frac{R_2}{R_1}}$, which means that X_2 does not provide any new information about S, i.e. X_2 is totally redundant given X_1. We observe in Fig. 2b that the MI related to X_1 (dotted) is tangent to the MI related to two modalities (plain) for $\rho = \sqrt{\frac{R_2}{R_1}}$, where the latter attains its minimum. In this case the optimal estimator of S upon X_1 and X_2, $\hat{S}_\rho(X_1,X_2)$, equals the optimal estimator based only on X_1, $\hat{S}(X_1)$ and the MMSE is maximum (Fig. 3b). Therefore, when $\rho = \sqrt{\frac{R_2}{R_1}}$ it is sufficient to use the modality X_1 alone. Otherwise, under the exact model assumption it is always beneficial to consider two modalities. Note that, for equal SNRs, i.e. $R_1 = R_2$, at $\rho = 1$, the two modalities are completely redundant since $X_1 = X_2$ (Fig. 2a and 3a).

B. Independent or correlated noises

In the previous subsection, we show that usually the MI related to two modalities is larger than the MI related to a unique modality. In this subsection, we wonder if the existence of links (correlation ρ) always guarantee higher MI. To this end, we compare the amount of information carried by both modalities when the link between noises exists ($\rho \neq 0$), $I_\rho\big(S;(X_1,X_2)\big)$ and when this link is inexistent ($\rho = 0$), $I_0\big(S;(X_1,X_2)\big)$. Following the Gaussian assumption, independence is equivalent to $\rho = 0$. Note that modalities with independent noises are not totally independent but conditionally independent given S. Using (10), we can write:

$$I_\rho\big(S;(X_1,X_2)\big) - I_0\big(S;(X_1,X_2)\big)$$
$$= I_\rho(S;X_2 \mid X_1) - I_0(S;X_2 \mid X_1). (12)$$

Eq (12) shows that modalities with independent noises outperform the ones with correlated noises when $I_\rho(S;X_2 \mid X_1) < I_0(S;X_2 \mid X_1)$. In Fig. 2b we can see that this occurs if $\rho \in \left]0, \rho_0 = \frac{2\sqrt{R_1R_2}}{R_1+R_2}\right[$, for which the noise correlation induces redundancy. When R_2 tends towards R_1, the modalities become more redundant and ρ_0 tends towards 1 (Fig. 2a). Conversely, when R_2 is negligible w.r.t R_1, ρ_0 tends to 0. Beyond this interval i.e $\rho \in [-1,0] \cup [\rho_0,1]$, $I_\rho(S;X_2 \mid X_1) \geq I_0(S;X_2 \mid X_1)$ since redundancy is exploited to enhance the estimation accuracy. In fact, the noise is reduced by summing the two modalities with proper scaling. In particular, I_ρ is infinite for $\rho = 1(R_1 \neq R_2)$ and $\rho = -1$ (Fig. 2b), since the noise is completely cancelled and S can be perfectly recovered (MMSE = 0, Fig. 3b) using respectively the optimal estimators (9):

$$\hat{S}_1(X_1,X_2) = \frac{\sqrt{R_1}X_1 - \sqrt{R_2}X_2}{\sqrt{R_1} - \sqrt{R_2}}. \quad (13)$$
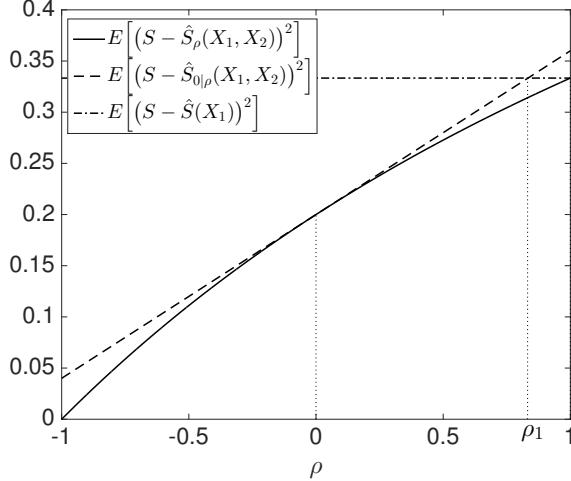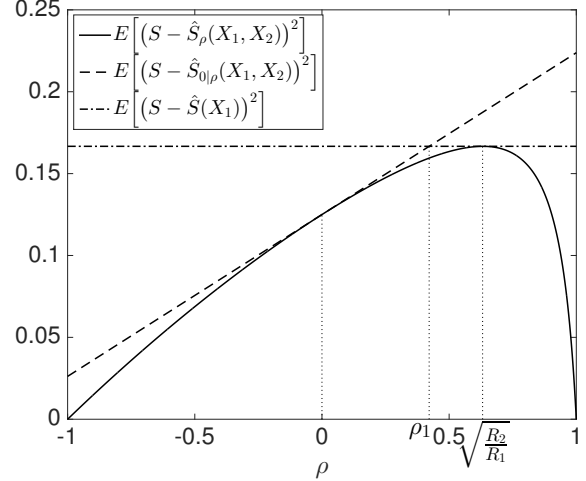
Fig. 3. Variation of the MSE as a function of ρ

$$\hat{S}_{-1}(X_1, X_2) = \frac{\sqrt{R_1}X_1 + \sqrt{R_2}X_2}{\sqrt{R_1} + \sqrt{R_2}}. \quad (14)$$

Notice that, for $R_1 = R_2$, the estimator (13) does not hold, because the two modalities are completely redundant (Fig. 2a). By contrast, for negatively-correlated noises, i.e $\rho \in [-1, 0[$, we always manage to reduce the noise whatever the values of the signal to noise ratios R_1 and R_2. Accordingly, we always get better results than the independent and positively correlated cases (Fig. 2b). Thus, to get the most out of the modalities, the noise correlation has to be negative. Otherwise, the noises have to be strongly correlated, i.e $\rho \in]\rho_0, 1]$, with different SNRs so that $\rho_0 \to 0$.

We can observe in figures 2 and 3 that the results of MI and MSE, although giving different points of view, are consistent which is not surprising since the gradient of the MI is directly related to the MSE [5] in Gaussian channels.

V. WRONG MODEL

In this section we study the loss of information when we wrongly define the noise correlation coefficient. We consider that ρ is unknown and denote by $\hat{\rho}$ the assumed noise correlation coefficient. To quantify the loss in information, we compare the information brought by the modalities considering the exact ρ, $I_\rho\big(S; (X_1, X_2)\big)$, with the "information" brought by the modalities about S when $\hat{\rho}$ is used instead of ρ, $I_{\hat{\rho}|\rho}\big(S; (X_1, X_2)\big)$ (remember it is not a MI cf. SectionIII).

A. $\hat{\rho} = 0$

Before investigating the loss for $\hat{\rho}$ in general, let us consider the case where $\hat{\rho} = 0$. In this particular case the modalities are assumed to have independent noises while the noises are actually correlated. In the literature, the noises between data sets are commonly assumed to be independent for computation simplicity. If the noises are actually correlated the error

induced by this assumption is given by (6). Using jointly Gaussian distributions, it follows:

$$I_\rho\big(S; (X_1, X_2)\big) - I_{0|\rho}\big(S; (X_1, X_2)\big) = \rho\frac{\sqrt{R_1 R_2}}{1 + R_1 + R_2}$$
$$+ \frac{1}{2}\log\frac{(1 + R_1)(1 + R_2) - (\sqrt{R_1 R_2} + \rho)^2}{(1 - \rho^2)(1 + R_1 + R_2)}. \quad (15)$$

For convenience, we denote this loss by $\Delta I(\rho, 0)$. The loss equals zero iff $\rho = 0$. Eq. (15) shows that the loss becomes infinite for $\rho = 1(R_1 \neq R_2)$ and $\rho = -1$. This is because S is not perfectly recovered with a wrong value $\hat{\rho} = 0$, while it is with the exact one. In fact, for infinite MI an exact quantification of the loss is given by the MSE. The MSE of estimating S using jointly X_1 and X_2 when the noises are assumed independent, is

$$E\left[\left(S - \hat{S}_{0|\rho}(X_1, X_2)\right)^2\right]$$
$$= \sigma^2\frac{1}{1 + R_1 + R_2} + 2\sigma^2\rho\frac{\sqrt{R_1 R_2}}{(1 + R_1 + R_2)^2}. \quad (16)$$

The error (16) is a linear function of ρ, and is tangent at $\rho = 0$ to the MSE of estimation using both modalities and exact noise correlation (Fig.3). Moreover, the error (16) intersects the MSE of estimation based uniquely of X_1 for $\rho = \rho_1$, where $\rho_1 = \frac{1 + R_1 + R_2}{2(1 + R_2)}\sqrt{\frac{R_1}{R_2}}$. Thus in the range $[\rho_1, 1]$, X_1 holds more information than does jointly X_1 and X_2 with the assumption of noises independence. Therefore, in this case using two modalities will give worse results for estimating S than using X_1 alone. Notice that if $R_2 \ll R_1$, ρ_1 tends towards zero. Furthermore, ρ_1 is never negative, which means that ignoring a negative correlation induces loss, but always insure more information than does the best (with the highest

SNR) single modality. Finally, note that $I_{0|\rho}\big(S;(X_1,X_2)\big)$ never reaches $I(S;X_2)$: it means that the MI of the joint observation, with wrong assumption $\hat\rho = 0$, is always better than using the worse single modality.

B. $\hat\rho \in [-1,1]$

In the general case, considering model (1) and assuming that $\hat\rho \neq \rho$ induces the following loss of information:

$$
\begin{aligned}
\Delta I(\rho,\hat\rho) \quad &(17)\\
= \frac{1}{2}\log &\frac{(1-\hat\rho^2)\big((1+R_1)(1+R_2)-(\sqrt{R_1R_2}+\rho)^2\big)}{(1-\rho^2)\big((1+R_1)(1+R_2)-(\sqrt{R_1R_2}+\hat\rho)^2\big)}\\
+ (\rho-\hat\rho) &\frac{(\sqrt{R_1}\hat\rho-\sqrt{R_2})(\sqrt{R_2}\hat\rho-\sqrt{R_1})}{(1-\hat\rho^2)\big((1+R_1)(1+R_2)-(\sqrt{R_1R_2}+\hat\rho)^2\big)}.
\end{aligned}
$$

The loss in (17) depends on the SNRs and the exact and wrong noise correlation coefficients, ρ and $\hat\rho$ respectively. Since, from (6), $\Delta I(\rho,\hat\rho)$ is the KL divergence between the exact and wrong densities for S given (X_1,X_2), it is always positive and equals zero iff $\hat\rho = \rho$. $\Delta I(\rho,\hat\rho)$ increases with $|\rho-\hat\rho|$, since the mismatch between the exact and wrong densities for S given (X_1,X_2) increases with the difference $|\rho-\hat\rho|$. Eq. (17) shows that $\Delta I(\rho,\hat\rho) \to +\infty$ for $\rho = 1 (R_1 \neq R_2)$ and $\rho = -1$, since the signal S can be perfectly recovered considering the exact noise correlation $(I_\rho\big(S;(X_1,X_2)\big) \to +\infty)$. Also, $\Delta I(\rho,\hat\rho) \to +\infty$ for $\hat\rho = 1(R_1 \neq R_2)$ and $\hat\rho = -1$, as we wrongly assume that the target signal could be perfectly recovered $(I_{\hat\rho|\rho}\big(S;(X_1,X_2)\big) \to -\infty)$. In the case of infinite information loss, an exact quantification of the loss is given by the MSE. The MSE of estimating S upon X_1 and X_2, considering the wrong noise correlation coefficient $\hat\rho$ is:

$$
\begin{aligned}
E\left[\Big(S - \hat S_{\hat\rho|\rho}(X_1,X_2)\Big)^2\right] \quad &(18)\\
= \sigma^2 &\frac{1-\hat\rho^2}{(1+R_1)(1+R_2)-(\sqrt{R_1R_2}+\hat\rho)^2}\\
+ 2\sigma^2(\rho-\hat\rho) &\frac{(\sqrt{R_1}\hat\rho-\sqrt{R_2})(\sqrt{R_2}\hat\rho-\sqrt{R_1})}{\Big((1+R_1)(1+R_2)-(\sqrt{R_1R_2}+\hat\rho)^2\Big)^2}.
\end{aligned}
$$

To investigate when the loss of information due to assuming $\hat\rho$ drops the information below that of one single modality, we compare $I_{\hat\rho|\rho}\big(S;(X_1,X_2)\big)$ with $I(S;X_i)$:

$$
\begin{aligned}
I_{\hat\rho|\rho}\big(S;(X_1,X_2)\big) - I(S;X_i) \quad\\
= I_\rho\big(S;X_j \mid X_i\big) - \Delta I(\rho,\hat\rho). \quad (19)
\end{aligned}
$$

Eq. (19) shows that, $I_{\hat\rho|\rho}\big(S;(X_1,X_2)\big) = I(S;X_i)$ when the loss of information induced by considering $\hat\rho \neq \rho$ cancels the additional information conveyed by X_j. For model (1), the equality $I_{\hat\rho|\rho}\big(S;(X_1,X_2)\big) = I(S;X_1)$ holds in two cases:

1) If $\hat\rho = \sqrt{\frac{R_2}{R_1}}, \forall\rho$

With this assumption, we consider that the modality X_2 is redundant w.r.t X_1 and ignore all the additional information given by X_2. In fact the optimal estimator of S using jointly X_1 and X_2 considering $\hat\rho$, $\hat S_{\hat\rho|\rho}(X_1,X_2)$ equals the optimal estimator of S using X_1 alone, $\hat S(X_1)$. This case can not occur for the second modality, according to the assumption $R_1 \geq R_2$.

2) If $\rho = \rho_1$ and $\hat\rho \neq \sqrt{\frac{R_2}{R_1}}$

where $\rho_1 = \hat\rho + \frac{(\sqrt{R_1}\hat\rho-\sqrt{R_2})\big((1+R_1)(1+R_2)-(\sqrt{R_1R_2}+\hat\rho)^2\big)}{2(1+R_1)(\sqrt{R_2}\hat\rho-\sqrt{R_1})}$. Unlike the first case, the optimal estimators using X_1 alone and using jointly X_1 and X_2 considering $\hat\rho$ are different *i.e.* $\hat S_{\hat\rho|\rho}(X_1,X_2) \neq \hat S(X_1)$. However, they have the same MSE *i.e.* $E\left[\big(S - \hat S_{\hat\rho|\rho}(X_1,X_2)\big)^2\right] = E\left[\big(S - \hat S(X_1)\big)^2\right]$. Which means that, X_1 alone and (X_1,X_2) jointly considering $\hat\rho$ communicate the same amount of information but different piece of information.

Finally, it is beneficial to use the best single modality X_1 alone if $\rho < \rho_1$ and $\sqrt{\frac{R_2}{R_1}} < \hat\rho$ or if $\rho > \rho_1$ and $\sqrt{\frac{R_2}{R_1}} > \hat\rho$ (since $I_{\hat\rho|\rho} < I(S;X_1)$). Furthermore, note that even the worse single modality X_2 can convey more information than does jointly X_1 and X_2 considering $\hat\rho$, for $\rho > \rho_2$. where:
$\rho_2 = \hat\rho + \frac{(\sqrt{R_2}\hat\rho-\sqrt{R_1})\big((1+R_1)(1+R_2)-(\sqrt{R_1R_2}+\hat\rho)^2\big)}{2(1+R_2)(\sqrt{R_1}\hat\rho-\sqrt{R_2})}$.

VI. CONCLUSION

In this paper, we raised the question of how the links between modalities can improve or decrease the estimation accuracy and information content. We considered a simple model and investigated the role of noise correlation which represents the link between two modalities. Our results suggests that the noise correlation usually increases information and boosts the performance. Moreover, negative noise correlation is the key to always get the most out of both modalities. However, under the assumption of a wrong noise correlation it may be beneficial to consider a single modality than two modalities. Future research should consider other characteristics of multimodality and more complex models.

REFERENCES

[1] D. Lahat, T. Adali, and C. Jutten, "Multimodal data fusion: An overview of methods, challenges, and prospects," *Proceedings of the IEEE*, vol. 103, no. 9, pp. 1449–1477, 2015.
[2] F. Peng and B. Chen, "Decentralized estimation with correlated additive noise: Does dependency always imply redundancy?" in *Signals, Systems and Computers, 2013 Asilomar Conference on*, Nov 2013, pp. 677–681.
[3] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. New York, NY, USA: Wiley-Interscience, 1991.
[4] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*. Upper Saddle River, NJ, USA: Prentice-Hall, Inc., 1993.
[5] D. Palomar and S. Verdu, "Gradient of mutual information in linear vector gaussian channels," *Information Theory, IEEE Transactions on*, vol. 52, no. 1, pp. 141–154, Jan 2006.
[6] S. Nirenberg, S. M. Carcieri, A. L. Jacobs, and P. E. Latham, "Retinal ganglion cells act largely as independent encoders," *Nature*, vol. 411, 2001. [Online]. Available: http://dx.doi.org/10.1038/35079612
[7] D. Lahat, T. Adali, and C. Jutten, "Challenges in multimodal data fusion," in *Signal Processing Conference (EUSIPCO), 2014 Proceedings of the 22nd European*, Sept 2014, pp. 101–105.