COHERENT ICA: IMPLICATIONS FOR AUDITORY SIGNAL PROCESSING

Simon Haykin

McMaster University
Hamilton, Ontario, Canada
haykin@mcmaster.ca

Kevin Kan

McMaster University
Hamilton, Ontario, Canada
kkan@soma.ece.mcmaster.ca

ABSTRACT

In this paper, we describe a novel algorithm, called Coherent Independent Components Analysis, and referred to as Coherent ICA for short. The algorithm, rooted in information-theoretic learning, exploits the combined use of the Infomax and Imax principles. Experimental results, based on the auditory coding of natural sounds, are presented that demonstrate the ability of coherent ICA to extract the envelope of amplitude-modulated sounds in a manner similar to the behaviour of neurons in the cochlear nucleus and inferior colliculus.

1. INTRODUCTION

Time manifests itself in many structural and functional specializations of the auditory system: With multiple time scales in acoustic stimuli, we find it informative to distinguish two specific components in the waveform of an acoustic stimulus [4]:

1. The *carrier*, represented by the fine structure of the waveform, which waxes and wanes in an "amplitude-modulated" fashion.
2. The *envelope*, which is the contour of the amplitude-modulated waveform.

From a physiological viewpoint, there is therefore interest in amplitude modulation, motivated by the desire to know whether envelope processing is actually embedded in the auditory system.

Indeed, across multiple layers of the auditory system, there are neurons that respond differently to an incoming amplitude-modulated speech signal. In particular, the successive layers of the auditory system distinguish themselves by responding to different limited ranges of amplitude-modulation rates: The lower layers are most responsive to fast changes in the energy of incoming acoustic stimuli, with progressively slower changes occurring in the higher layers.

In light of this reality, it is not surprising that amplitude modulation is considered to be an important acoustic cue in the perception of sound, and may therefore play an equally significant role in the design of a cocktail party processor [3].

With auditory processing as the issue of interest, the first question that we address in this paper is the following:

> **Given an additive mixture of amplitude-modulated speech signals, how can we separate the envelopes of the individual components, ignoring the associated carriers?**

Another important question addressed in the paper is:

> **In a self-organized manner, can we learn the manner in which the different processing layers in the auditory system respond to an amplitude-modulated stimulus?**

The answers to these basic two questions are to be found in a new learning principle termed "coherent independent components analysis", which, henceforth, is referred to simply as coherent ICA [5]. The formulation of coherent ICA is rooted in information-theoretic learning, as described next.

2. COHERENT ICA

The maximization of mutual information principle, commonly known as the *Infomax principle* [6], stands out in a dominant way in the formulation of information-theoretic learning models. The Infomax principle not only plays a significant role in our understanding of redundancy reduction, the modeling of perception, and the extraction of independent components [2], but also its variant, the *Imax principle*, due to Becker [1], plays an important role of its own in the extraction of spatially coherent features. In reality, Infomax and Imax play complementary roles in the following sense:

> **Infomax deals with information flow across a network, whereas Imax deals with spatial coherence across a pair of network outputs.**

In the literature, Infomax and Imax have been treated as two unrelated principles, ignoring their complementary roles. In this paper, starting with the two-network structure of Figure 1, we take the opposite view by exploiting their combined use, which leads to the

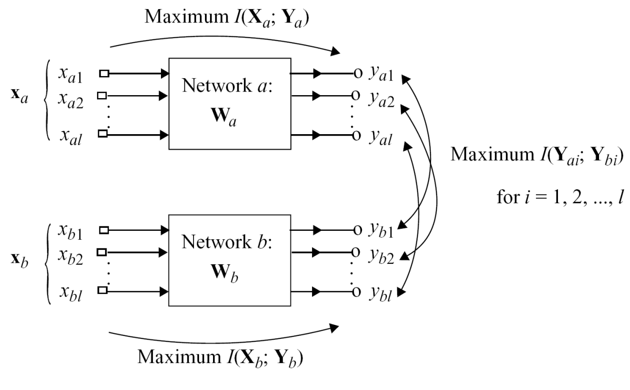
Maximum $I(\mathbf{X}_a; \mathbf{Y}_a)$
$\mathbf{x}_a$
x_{a1}
x_{a2}
x_{al}
Network a: $\mathbf{W}_a$
y_{a1}
y_{a2}
y_{al}
Maximum $I(\mathbf{Y}_{ai}; \mathbf{Y}_{bi})$
for $i = 1, 2, ..., l$
$\mathbf{x}_b$
x_{b1}
x_{b2}
x_{bl}
Network b: $\mathbf{W}_b$
y_{b1}
y_{b2}
y_{bl}
Maximum $I(\mathbf{X}_b; \mathbf{Y}_b)$

Figure 1: The coupled-network layout for coherent ICA.

formulation of coherent ICA, hence the reference to it as a "coherent" signal-processing algorithm.

To elaborate on the scenario depicted in Figure 1, where we have two separate but dimensionally similar neural networks; network a is characterized by the weight matrix $\mathbf{W}_a$, and the other network b is characterized by the weight matrix $\mathbf{W}_b$.The goal is to combine the principles of Infomax and Imax into a new learning strategy for the self-organized training of these two networks. In particular, the strategy is to be configured in such a way that the two aforementioned properties, information flow in each network in accordance with the Infomax principle and spatial coherence across the neuronal outputs of the two networks on a *pair-by-pair basis* in accordance with the Imax principle, are integrated into a composite learning principle.

Consider, first, the Infomax principle applied across the input-output of each network in Figure 1. Network a is characterized by the mutual information

$$I(\mathbf{X}_a;\mathbf{Y}_a) = \mathbf{E}[\text{log}p_{\mathbf{Y}_a}(\mathbf{y}_a)] \qquad (1)$$

where we have ignored an additive constant that is independent of $\mathbf{W}_a$ and therefore immaterial; $\mathbf{E}$ is the expectation operator. With the elements of the output vector $\mathbf{Y}_a$ being statistically independent in accordance with ICA, we may express the probability density function (pdf) of $\mathbf{Y}_a$ as

$$p_{\mathbf{Y}_a}(\mathbf{y}_a) = \prod_{i=1}^{l} p_{Y_{ai}}(y_{ai})$$

and therefore go on to rewrite (1) in the equivalent form

$$I(\mathbf{Y}_a;\mathbf{X}_a) = \mathbf{E}\left[\sum_{i=1}^{l} \text{log}p_{Y_{ai}}(y_{ai})\right] \qquad (2)$$

Similarly, for the other network b, we have

$$I(\mathbf{X}_b;\mathbf{Y}_b) = \mathbf{E}\left[\sum_{i=1}^{l} \text{log}p_{Y_{bi}}(y_{bi})\right] \qquad (3)$$

Consider next the Imax principle applied across the output terminals of the two networks, treated on a pair-by-pair basis; we may express the mutual information between the outputs $\mathbf{Y}_{ai}$ and $\mathbf{Y}_{bi}$ in terms of the copula[1] as

$$I(\mathbf{Y}_{ai};\mathbf{Y}_{bi}) = \mathbf{E}\left[\sum_{i=1}^{l} \text{log}C_{Y_{ai},Y_{bi}}(y_{ai},y_{bi})\right], \quad i=1,2,\ldots,l \qquad (4)$$

Here, again, with the l outputs of each network in Figure 1 assumed to be statistically independent, the individual contributions in (4) are additive, yielding the sum

$$\sum_{i=1}^{l} I(Y_{ai};Y_{bi}) = \mathbf{E}\left[\sum_{i=1}^{l} \text{log}C_{Y_{ai},Y_{bi}}(y_{ai},y_{bi})\right] \qquad (5)$$

To combine the contributions described in (2),(3) and (5) into a single ensemble-averaged *objective function* that accounts for the applications of the Infomax and Imax principles, we simply write[2] (after the combination and simplification of terms)

$$J(\mathbf{W}_a,\mathbf{W}_b) = \mathbf{E}\left[\sum_{i=1}^{l} \text{log}p_{Y_{ai},Y_{bi}}(y_{ai},y_{bi}))\right] \qquad (6)$$

where $p_{Y_{ai},Y_{bi}}(y_{ai},y_{bi})$ is the joint probability density function of the random variables Y_{ai} and Y_{bi} represented by the sample values of the network output, y_{ai} and y_{bi}, respectively for all $i = 1,2,\ldots,l$. We may now make the statement:

The coherent ICA principle maximizes the joint objective function $J(\mathbf{W}_a,\mathbf{W}_b)$ with respect to the weight matrices $\mathbf{W}_a$ and $\mathbf{W}_b$.

Let $\mathbf{w}_{ai}^T$ and $\mathbf{w}_{bi}^T$ denote the ith row vectors of the weight matrices $\mathbf{W}_a$ and $\mathbf{W}_b$, respectively. We may then express

$$\begin{aligned}\mathbf{y}_i &= \begin{bmatrix} y_{ai} \\ y_{bi} \end{bmatrix} \\ &= \begin{bmatrix} \mathbf{w}_{ai}^T\mathbf{x}_{ai} \\ \mathbf{w}_{bi}^T\mathbf{x}_{bi} \end{bmatrix}, \quad i=1,2,\ldots,l \qquad (7)\end{aligned}$$

Typically, the data, drawn from natural scenes, tend to be *sparse*. To satisfy this property, we take the distributions of the composite output vector $\mathbf{y}_i$ to include a zero-mean generalized Gaussian bivariate distribution with a two-by-two covariance matrix $\mathbf{\Sigma}$, as shown by

$$p_{\mathbf{Y}_i}(\mathbf{y}_i) = \frac{1}{2\pi(\text{det}\mathbf{\Sigma})^{1/2}}\exp\left(-\frac{1}{2}(\mathbf{y}_i^T\mathbf{\Sigma}^{-1}\mathbf{y}_i)^{\alpha/2}\right), \quad i=1,2,\ldots,l \qquad (8)$$

where the parameter α controls the shape and sparseness of the copula. The covariance matrix $\mathbf{\Sigma}$ is itself defined by

$$\mathbf{\Sigma} = \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix} \qquad (9)$$

[1] According to *Sklars theorem* on copulas [8]:
Given the cumulative distribution functions $P_{X,Y}(x,y)$, $P_X(x)$, and $P_Y(y)$, pertaining to the random variables X and Y, there exists a unique copula $C_{U,V}(u,v)$ that satisfies the following pair of relationships:

$$P_{X,Y}(x,y) = C(P_X(x),P_Y(y))$$

and

$$C_{U,V}(u,v) = P(P_X^{-1}(x),P_Y^{-1}(y))$$

where the two new random variables U and V are respectively defined by the nonlinear transformations of X and Y; that is,

$$U = P_X(x)$$

and

$$V = P_Y(y)$$

[2] To be rigorous, in place of (6) we should write

$$J(\mathbf{W}_a,\mathbf{W}_b) = \mathbf{E}\left[\sum_{i=1}^{l}\text{log}p_{Y_{ai}}(y_{ai}) + \sum_{i=1}^{l}\text{log}p_{Y_{bi}}(y_{bi})\right] + \lambda\sum_{i=1}^{l}\mathbf{E}[\text{log}C_{Y_{ai},Y_{bi}}(y_{ai},y_{bi})]$$

where the parameter λ balances the contributions of Infomax and Imax. In (6), we have set λ = 1 to simplify the formulation of coherent ICA. From a theoretical and practical perspective, the generalization of coherent ICA through the inclusion of the control parameter λ warrants investigation. However, as we will see later on, in the absence of λ we are still able to exercise a trade-off between the contributions of Infomax and Imax through a correlation coefficient ρ, yet to be defined.

where the *correlation coefficient* ρ controls the extent of correlation between the paired outputs y_{ai} and y_{bi} for $i = 1, 2, \ldots, l$. Increasing ρ affects the relative importance of Imax over Infomax by favouring a more "coherent process" being learned jointly by the two networks. Substituting (7) through (9) into (6) and ignoring the constant term $2\pi(\det\mathbf{\Sigma})^{1/2}$ for some prescribed ρ, we get the reformulated objective function

$$J(\mathbf{W}_a, \mathbf{W}_b) = -\frac{1}{2}\mathbf{E}\left[\sum_{i=1}^{l}(\mathbf{y}_i^T\mathbf{\Sigma}^{-1}\mathbf{y}_i)^{\alpha/2}\right] \qquad (10)$$

where the ensemble averaging is performed with respect to the $\mathbf{y}_i$'s. With this objective function at hand, we may now derive an algorithmic implementation of the coherent ICA principle by using the instantaneous values of the quadratic term $\mathbf{y}_i^T\mathbf{\Sigma}^{-1}\mathbf{y}_i$ for all i, as an estimate of the expectations in (10), thereby obtaining

$$\begin{aligned}\hat{J}(\mathbf{W}_a, \mathbf{W}_b) &= -\frac{1}{2}\sum_{i=1}^{l}(\mathbf{y}_i^T\mathbf{\Sigma}^{-1}\mathbf{y}_i)^{\alpha/2}\\ &= -\frac{1}{2(1-\rho^2)}\sum_{i=1}^{l}(y_{ai}^2 - 2\rho y_{ai}y_{bi} + y_{bi}^2)^{\alpha/2}\end{aligned} \qquad (11)$$

where we have used the hat in $\hat{J}(\mathbf{W}_a, \mathbf{W}_b)$ to express it as an estimate of the ensemble-averaged objective function $J(\mathbf{W}_a, \mathbf{W}_b)$.

To obtain the adaption rule for the weighted vector $\mathbf{w}_{ai}$, we use the chain rule of calculus to write

$$\begin{aligned}\frac{\partial\hat{J}(\mathbf{W}_a, \mathbf{W}_b)}{\partial\mathbf{w}_{ai}} &= \frac{\partial\hat{J}(\mathbf{W}_a, \mathbf{W}_b)}{\partial y_{ai}}\frac{\partial y_{ai}}{\partial\mathbf{w}_{ai}}\\ &= -\frac{\alpha}{2(1-\rho^2)}(y_{ai} - \rho y_{bi})(y_{ai}^2 - 2\rho y_{ai}y_{bi} + y_{bi}^2)^{\frac{\alpha}{2}-1}\mathbf{x}_a\end{aligned} \qquad (12)$$

where it is assumed that the variance of both outputs is unity. Accordingly, the adjustment applied to the weight vector $\mathbf{w}_{ai}$ is defined by

$$\begin{aligned}\Delta\mathbf{w}_{ai} &= -2\eta\frac{\partial\hat{J}(\mathbf{W}_a, \mathbf{W}_b)}{\partial\mathbf{w}_{ai}}\\ &= \frac{\alpha\eta}{1-\rho^2}(y_{ai} - \rho y_{bi})(y_{ai}^2 - 2\rho y_{ai}y_{bi} + y_{bi}^2)^{\frac{\alpha}{2}-1}\mathbf{x}_a\end{aligned}$$

$$\text{for all } i \qquad (13)$$

where the learning-rate parameter η is assumed to be common to both networks, and $i = 1, 2, \ldots, l$. Similarly, the adjustment applied to the weight vector $\mathbf{w}_{bi}$ is defined by

$$\Delta\mathbf{w}_{bi} = \frac{\alpha\eta}{1-\rho^2}(y_{bi} - \rho y_{ai})(y_{ai}^2 - 2\rho y_{ai}y_{bi} + y_{bi}^2)^{\frac{\alpha}{2}-1}\mathbf{x}_b$$

$$\text{for all } i \qquad (14)$$

During the learning process, it is assumed that the inputs $\mathbf{x}_a$ and $\mathbf{x}_a$ in Figure 1 are both *whitened* prior to processing; this is normal practice in ICA-related work. Moreover, after each iteration of the learning process, the weights are *normalized*, which constrains the variance of the network outputs to unity. We may thus express the weight update applied to network a as

$$\mathbf{w}_{ai} \leftarrow \mathbf{w}_{ai} + \Delta\mathbf{w}_{ai} \qquad (15)$$

$$\mathbf{w}_{ai} \leftarrow \frac{\mathbf{w}_{ai}}{\parallel\mathbf{w}_{ai}\parallel} \qquad (16)$$

for all i; and similarly for network b.

For applications that involve the modeling of data where we have two streams consisting of spatially shifted data, as that described in Figure 1, it is useful to enforce a *weight-sharing* constraint between the two streams, in which case we set

$$\mathbf{w}_{ai} = \mathbf{w}_{bi} \quad \text{for all } i$$

Thus, by starting the weight-adaptation rule for coherent ICA with the same initial weight-matrices assigned to networks a and b, the weight sharing is maintained at every step of the adaptation rule.

3. APPLICATION OF COHERENT ICA TO AUDITORY CODING OF NATURAL SOUNDS

In coherent ICA, the goal is to extract information that is maintained "coherent" across separate sources while, at the same time, information flow across the networks associated with the sources is maximized. Since in amplitude modulation, the envelope varies slowly compared to the carrier, we may view it as a form of temporal coherence in a limited sense; that is, across two time-steps Δt seconds apart, we may set $x(t + \Delta t) \approx x(t)$.

For an illustrative application of coherent ICA, we applied it to a set of speech samples of English speakers taken from the TIMIT database. The set of sounds comprised an equal number of male and female speakers, with a total duration of approximately five minutes. The data were first passed through a grammatone filter using the Auditory tool Box due to Slaney [9]. The filter bandwidths were set equal to the equivalent rectangular bandwidth measurements in humans. In the experiment reported herein, a single filter was used, centered at 2 kHz. The resulting output data were then half-wave rectified (HWR), corresponding to the primary nonlinearity of the inner-hair cells, yielding the signal

$$x(t) = \text{HWR}(h_i(t) * x_{\text{input}}) \qquad (17)$$

where $h_i(t)$ is the impulse response of the grammatone filter, x_{input} is its input, and '$*$' denotes convolution.

In applying the coherent ICA model to the data set, the problem can be approached in different ways, depending on how the input streams applied to networks a and b in Figure 1 are chosen. For our experiment, we chose to mimic the *slow-feature analysis model*, where coherence is maximized between two successive time-steps [10]. The two inputs were overlapping, and in order to prevent the weights from converging onto trivial solutions, the weight-sharing constraint was enforced. Thus, we set the weight matrices

$$\mathbf{w}_{ai} = \mathbf{w}_{bi} \quad \text{for all } i$$

and if the input applied to network a is

$$\mathbf{x}_a(t) = \mathbf{x}(t)$$

then the input applied to network b is

$$\mathbf{x}_b(t) = \mathbf{x}(t + \Delta t)$$

where Δt is the duration between two successive time-steps.

The coherent ICA algorithm was applied twice to the data to learn two successive layers of filters. The first layer used the

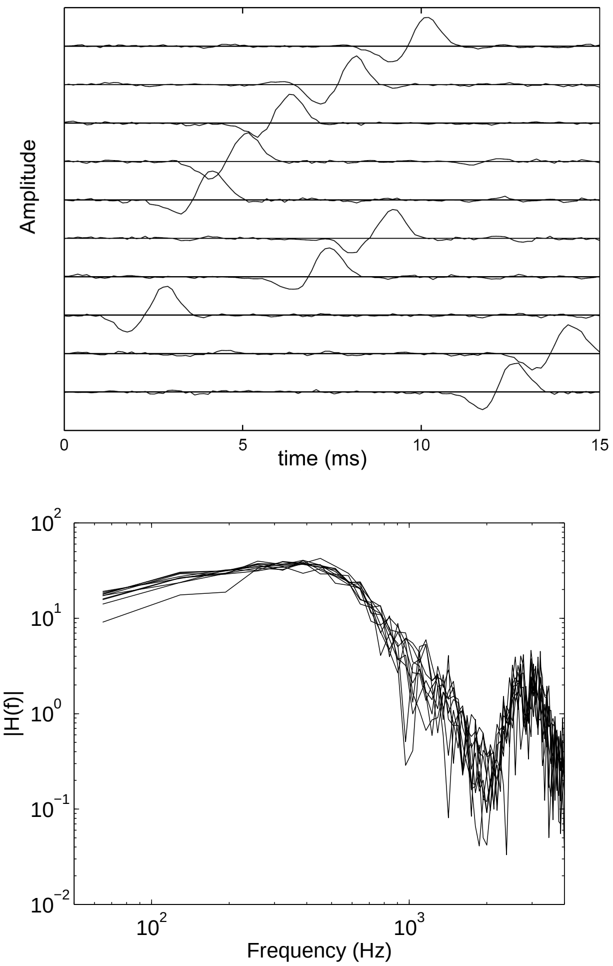

Figure 2: Time and frequency plots obtained for the first layer of ten filters learned in the experiment on auditory coding of natural sounds, using coherent ICA.

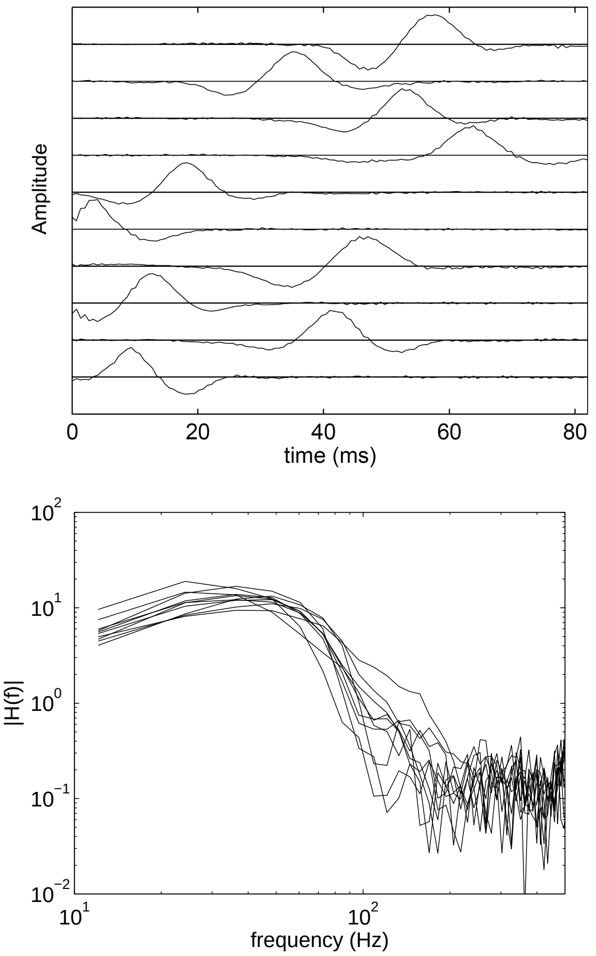

Figure 3: Time and frequency plots obtained for the second layer of another set of ten filters learned in the experiment on auditory coding of natural sounds, using coherent ICA.

filtered and rectified speech signal $x(t)$ of (17) as input. After the ten filters were learned, $x(t)$ was convolved with one of them and then half-wave rectified as before to constrain the output to remain above zero. The resulting output was downsampled from 8 kHz to 1 kHz. Then, a second set of ten filters was learned by treating the processed output of the first layer as the input to the second layer. Structurally, the model just described is identical to the phenomenological model proposed by Nelson and Carney [7], with one key difference:

> **The filters in our experiment were learned using coherent ICA, whereas in the Nelson-Carney model the filters were tuned using physiologically plausible data to obtain a specific result.**

Figures 2 and 3 show the filters learned using coherent ICA applied to the speech data for the first and second layers of auditory processing. All the filters are found to be smooth and temporally localized. Most importantly, the figures clearly show two important features:

1. The passband of the filters learned in both layers only includes frequencies within the modulation spectrum, ignoring the carriers altogether.
2. The baseband filters learned in the first layer of processing have a cutoff frequency that is about ten times that of the baseband filters learned in the second layer of processing. In other words, the first layer of our experimental model (based on coherent ICA) is most responsive to fast changes in the input auditory signal whereas the second layer of the model is responsive to slower changes in the input. This result is exactly what we alluded to in the Introduction.

In short, the two set of filters learned by coherent ICA, applied to natural sounds, are baseband (modulation) filters that appear to exhibit properties similar to those of biological neurons in the cochlear nucleus and inferior colloculus.

4. CONCLUDING REMARKS

In this paper, we have presented some novel ideas, summarized as follows:

1. We drew attention to the notion of copulas in statistics, which can be used to describe the dependencies between random variables without regard to marginal distributions. (Copulas do not seem to be known in the signal processing and information theory literature).
2. We exploited the combined use of the Infomax and Imax principles to formulate the new coherent ICA principle, which was implemented in algorithmic form.
3. We used the coherent ICA algorithm to explain the role of amplitude modulation in auditory processing in the brain, which was done by applying it to a database of natural sounds. In particular, we addressed the two basic questions posed in the introduction experimentally, by presenting two important results:
   (i) The ability of coherent ICA to exhibit amplitude-modulation tuning, thereby supporting the notion that envelope processing is embedded in the auditory system.
   (ii) The ability of coherent ICA to learn the varying rates at which two successive processing layers of filters respond to acoustic stimuli in a manner that mimics what goes on in the auditory system.

In describing the coherent ICA algorithm and experimenting with it , we have set the stage for future research in several new directions:

- Further work on neurobiological plausibility of the algorithm,
- Algorithmic refinements, and
- Novel auditory signal-processing applications of coherent ICA, including the cocktail party processor.

Acknowledgement

The authors are grateful to the Natural Sciences and Engineering Research council of Canada for providing the financial support to do the work described in this paper.

5. REFERENCES

[1] S. Becker (1992), *An Information-theoretic Unsupervised Learning Algorithm for Neural Networks*, University of Toronto, Ontario, Canada.

[2] S. Haykin, *Neural Networks: A Comprehensive Foundation*, 2nd edition, Prentice-Hall, 1999.

[3] S. Haykin and Z. Chen (2007), *The Machine Cocktail Party Problem*. In S. Haykin, J. Principe, T. Sejnowski, and J. McWhirter, *New Directions in Statistical Signal Processing: From Systems to Brains*, MIT Press.

[4] P.X. Joris, C.E. Schreiner, and A. Rees (2004), “Neural Processing of Amplitude-modulated Sounds”, *Physical Review*, 84: 541-577.

[5] K. Kan (2007), *Coherent Independent Component Analysis: Theory and Applications,* Masters Thesis, McMaster University, Ontario, Canada.

[6] R. Linsker (1988), “Self-organization in a Perceptual Network”, *Computer*, 21: 105-117.

[7] P.C. Nelson and L.H. Carney (2004), “A Phenomenological Model of Peripheral and Central Neural Responses to Amplitude-modulated Tones”, *Journal of the Acoustical Society of America*, 116(4): 2172-2186.

[8] A. Sklar (1959), “Fonctions de répartition à n dimensions et leurs marges”, Publications de l’Institut de Statistique de L’Université de Paris 8, 229-231.

[9] M. Slaney (1998), *Auditory Box: A Matlab Toolbox for Auditory Modeling Work*, Technical Report TR1998-010, Interval Research Corporation.

[10] J. Wiskott and T.J. Sejnowski (2002), “Slow Feature Analysis: Unsupervised Learning of Invariance”, *Neural Computation*, 14: 715-770.