Evaluating Information Transfer Between Auditory Cortical Neurons

Boris Gourevitch and Jos J. Eggermont

Department of Physiology and Biophysics and Department of Psychology, University of Calgary, Calgary, Alberta, Canada

Submitted 16 October 2006; accepted in final form 21 December 2006

INTRODUCTION

Over the past decades, most investigations concerning spike trains have focused on one of the main issues in systems neuroscience—understanding of the neural code (deCharms and Zador 2000; Eggermont 1998; Perkel and Bullock 1968; Pouget et al. 2000). Basically, the two main approaches depict neurons either as sensitive to input firing rates and act as integrators or as coincident detectors that are sensitive to the temporal patterns of the input (Gerstein and Kirkland 2001). Given that the synchronization between the two multiple single-unit recordings has a modest dependency on the sensory stimulus used (Eggermont 1994), coincident firing has mainly been investigated and used in the search for neural assemblies or neuron clusters able to show a stimulus-induced modulation of their synchronized firing [e.g., the gravitational clustering method (Baker and Gerstein 2000; Gerstein and Aertsen 1985) and synchrony clustering (Eggermont 2006)]. Studies based on firing rates have instead mainly focused on the processing abilities of single neurons as a consequence of their unique physiological specialization (type of cell, tuning properties), how they integrate the input activity, and eventually how a population of neurons possibly encodes one feature of the stimulus by a rate code (Pouget et al. 2000). Consequently, there has been a steady rise in interest in the estimation of “information” carried by single neuron, and multiple single units or populations to specific stimuli.

The concept of information with respect to neuronal activity, although stemming from Shannon’s theory, has no standardized meaning in neuroscience and has been used in several different ways (Borst and Theunissen 1999). For instance, information rates have been estimated at a microscopic physiological scale in the transmembrane response to hormonal stimuli (Prank et al. 2000) or within a single synapse (London et al. 2002). The entropy carried by a spike train has also been widely investigated, the complexity of estimation being reflected by the large range of methods proposed [e.g., histogram method (Strong et al. 1998), vector spaces (Vicario 2002), Lempel–Ziv complexity (Amigo et al. 2004)]. Information-related measures between neurons have also been used to study the effect of noise correlation on encoding and decoding stimuli (Averbeck and Lee 2006). Finally, the mutual information between the spike-train responses and a set of stimuli was also estimated to investigate the discrimination abilities of neurons (Borst and Theunissen 1999; Chechik et al. 2006; Gehr et al. 2000; Werner and Mountcastle 1965).

Herein we present the transfer entropy as a new exploratory tool that provides a bridge between the study of neural assemblies and the information about stimuli carried by individual neurons. More precisely, the transfer entropy estimates the part of activity of a neuron that is not dependent on its own past but dependent on the past activity of another neuron. In a nutshell, it estimates the information transferred between two neurons in both directions. To our knowledge, the transfer entropy concept has been discussed in Jumarie (1990), but applied only once in context of physical continuous systems (Schreiber 2000), and has never been applied to spike trains. Yet, to a certain extent, this tool is able to distinguish information resulting from common history and exclude it by appropriate conditioning of the entropy. Transfer entropy also detects asymmetry in neural relations, allowing studies of possible feedback in neural circuits, a topic that recently gained considerable interest (Contreras et al. 1996; Hupe et al. 1998; Krupa et al. 1999; Sillito et al. 1993, 1994; Yan and Suga 1996). Finally, but not unimportantly, transfer entropy takes into account linear and nonlinear flows and thus may represent a very general way to define the causality strength between two spikes trains. In particular, the window size for which maximum information is transferred may be useful to study neural integrative properties.

After presenting the mathematical tenets of the statistics, its basic properties will be elucidated through simulations of independent Poisson processes and compared with those of cross-correlation measures. An exploration of the multiunit...
activity recorded with an array of 16 electrodes in cat auditory cortex will highlight the potential of the method in network studies. Finally, recordings of spontaneous activity in 21 cats will show the statistical distribution of transfer entropy and above all delineate the size of temporal integration windows in primary auditory cortex as a potential first important physiological result.

METHODS

Transfer entropy

Let \( X_1 \) and \( X_2 \) be two spike trains. Let \( X_1^f(t), X_2^f(t) \) be the number of spikes of \( X_1 \) and \( X_2 \), respectively, falling in the upcoming time interval \( [t, t + \tau_p] \). Similarly, let \( X_1^p(t), X_2^p(t) \) be the number of spikes of \( X_1^p \) and \( X_2^p \) falling in the past time interval \( [t - \tau_f, t] \), and \( \tau_f \) and \( \tau_p \), typically range from 1 to 100 ms. In practice, \( \tau_f \) is considered discrete with \( \tau_f = n \tau_f \), \( n \in \{ 0, 1, 2, \ldots \} \) such that \( \{ X_1^f(t_n) \}, \{ X_2^f(t_n) \}, \{ X_1^p(t_m) \}, \{ X_2^p(t_m) \} \) are discrete processes.

In the stationary case, the transfer entropy from \( X_1 \) to \( X_2 \) can be defined as the amount of mutual information between the past of \( X_1 \) (\( X_1^f \)) and the future of \( X_2 \) (\( X_2^p \)) when the past of \( X_2 \) (\( X_2^f \)) is already known, i.e.

\[
TE_{X_1 \rightarrow X_2} = h(X_2^f | X_1^f) - h(X_2^f | X_1^f, X_2^p)
\]

where \( h(X_2^f | X_1^f) \) is the entropy of the process \( X_2^f \) conditional on its past. The distributions of \( X_1^f, X_2^f, X_1^p, X_2^p \), being discrete, can be written explicitly as

\[
TE_{X_1 \rightarrow X_2} = h(X_2^f | X_1^f) - h(X_2^f | X_1^f, X_2^p) = \sum_k \sum_i \sum_j P(X_1^f = k, X_2^f = j) \times \log [\frac{P(X_2^f = j | X_1^f = k)}{P(X_2^f = j)}]
\]

where \( k, j, l \in \{ 0, 1, 2, \ldots \} \). Under independence between \( X_1 \) and \( X_2 \), \( TE_{X_1 \rightarrow X_2} = 0 \). The equality

\[
TE_{X_1 \rightarrow X_2} = h(X_2^f | X_1^f) - h(X_2^f | X_1^f, X_2^p)
\]

shows that the transfer entropy (TE) represents the amount of information provided by the additional knowledge of the past of \( X_1 \) in the model describing the information between the past and the future of \( X_2 \).

When no hypothesis regarding the distribution of \( X_{1/2}^{fp/p} \) is made, the theoretical properties of transfer entropy are probably extremely difficult to know a priori. Nevertheless, if \( \tau_f \) and \( \tau_p \) are large, the joint distribution between \( X_1^f, X_2^p \), and \( X_1^p, X_2^f \) will be broad and sparse. The transfer entropy will thus be increasing automatically even if there is no causal link. Two corrections are necessary to avoid this effect. We present them in the next several paragraphs.

Bias

At first, we remove from \( TE_{X_1 \rightarrow X_2} \) an estimate of the same transfer entropy in shuffled data, thereby modifying \( X_1^f \) to make it independent of \( X_2^{fp/p} \). Practically, we randomly shuffle the interspike intervals (ISIs) of \( X_1 \), which does not change the ISI distribution of \( X_1 \) but completely disconnects \( X_1^f \) and \( X_2^{fp/p} \). This procedure was previously used in another context in Hung et al. (2002), for instance. Another possibility is to directly shuffle the values of the process \( X_1^f \), which gives similar results and may be faster computationally. The shuffled estimate is dubbed \( TE_{X_1 \rightarrow X_2}^{shuffled} = I_{X_1^f, X_2^p}^{shuffled} \) in the following and is an average of results obtained on \( n \) trials.

Normalization

Finally, given the properties of mutual information, we define the normalized transfer entropy (NTE) by

\[
NTE_{X_1 \rightarrow X_2} = \frac{TE_{X_1 \rightarrow X_2} - TE_{X_1 \rightarrow X_2}^{shuffled}}{h(X_2^f)} \in [0, 1]
\]

Intuitively, this represents the fraction of information in \( X_2 \) not explained by its own past and explained by the past of \( X_1 \).

Preferred direction of flow

Similarly to Wang and Kadia (2001) and Schnupp et al. (2006), we also define a selective direction of the preferred direction of flow (DF) by

\[
DF_{X_1 \rightarrow X_2} = \frac{NTE_{X_1 \rightarrow X_2} - NTE_{X_2 \rightarrow X_1}}{NTE_{X_1 \rightarrow X_2} + NTE_{X_2 \rightarrow X_1}} \in [-1, 1]
\]

where, obviously, \( DF_{X_1 \rightarrow X_2} = -DF_{X_2 \rightarrow X_1} \).

Final estimates

All previous quantities depend on \( \tau_f \) and \( \tau_p \) values. All simulations and investigations on real data convinced us that a clear peak always exists in the surface of NTE values as a function of \( \tau_f \) and \( \tau_p \). Consequently, we always assign to NTE the maximum over the set of \( \tau_f \) and \( \tau_p \) values.

Cross-correlation

Cross-correlograms were calculated using custom-made programs in MATLAB (Eggermont and Smith 1995b). The bin size was 2 ms and the resulting cross-correlogram was smoothed with a three-bin running average. Stationarity estimates of the recordings were based on firing rate (mean and variance) in 100-s-long segments of the 15-min recordings for silence. To correct for the overall firing rate, burst firing, and common periodicities in the firing of the neurons, the cross-covariance was deconvolved with the square root of the product of the autocovariance functions. This deconvolution was done in the frequency domain, where it becomes a simple division; Fourier transformation back to the time domain resulted in the corrected cross-correlation coefficient function.

Simulations

We simulated some independent Poisson processes for \( X_1 \) and \( X_2 \). We then replaced in \( X_2 \) a proportion \( \alpha \in [0, 1] \) of spikes by the same proportion of spikes of \( X_1 \) delayed by 10 ms. In this way, the firing rate (FR) of \( X_2 \) is not modified, but it creates a causality link from \( X_1 \) to \( X_2 \) of various strengths proportional to \( \alpha \). The parameter for the exponential distribution underlying the Poisson process is \( \lambda = 1/10 \), which gives an average FR of 10 spikes/s for \( X_1 \) and \( X_2 \). Spike trains of 300-s duration are generated for both processes.

Real data

We analyzed spike trains recorded during silence from the primary auditory cortex (AI) of 21 ketamine-anesthetized cats. The length of the recording is 900 s for each data set. Recordings were made with two arrays of eight microelectrodes, arranged in a 4 × 2 pattern with 0.5-mm separation between electrodes. The arrays were independently inserted into the auditory cortex. Details about the anesthesia, the electrode array, and the protocol can be found in Tomita and Eggermont (2005). The spike trains from individual electrodes repre-
Simulations were used. The reason that contrast to the linear dependency for the cross-correlation close to zero.

Results

To compute each value of shuffled estimates \( n = 20 \) trials were used.

Simulations

For \( \alpha = 1 \) (full causality, Fig. 1A), \( \text{NTE}_{X_{1}\rightarrow X_{2}} = 1 \) for \( \tau_{p} = \tau_{f} = 10 \) ms and \( \text{NTE}_{X_{2}\rightarrow X_{1}} = 0 \) (Fig. 1B) as is discussed in the following. The reason that \( \text{NTE} \) reaches its maximal value for \( \tau_{p} = \tau_{f} = 10 \) ms is explained through the example in Fig. 2, which also explains why we find \( \text{NTE}_{X_{1}\rightarrow X_{2}} = 0 \) for \( \tau_{p} + \tau_{f} < 10 \) ms.

The peak in Fig. 1C, corresponding to the maximum of transfer entropy, is less sharp than that for the \( \text{NTE} \) estimate. The preferred direction of flow (Fig. 1D) is always \( \text{DF}_{X_{1}\rightarrow X_{2}} = 1 \) except when \( \tau_{p} + \tau_{f} < 10 \) ms, where \( \text{NTE}_{X_{1}\rightarrow X_{2}} = 0 \) as explained previously. Consequently, the \( \text{DF} \) statistic should not be used when both \( \text{NTE} \) estimates are close to zero.

The \( \text{NTE} \) estimate is nonlinearly related to \( \alpha \) (Fig. 3), in contrast to the linear dependency for the cross-correlation (XC). However, \( \text{NTE} \) is more suited to the study of complex neural networks than XC.

Figure 4A, model 1 represents the case \( \alpha = 0.6 \) with a delay of 10 ms between \( X_{1} \) and \( X_{2} \). Model 2 (Fig. 4B) is a combination of the same 60% of spikes of \( X_{1} \) but in three fractions of 20%, each part being delayed by 4, 8, and 10 ms, respectively. Such variability in delays may occur if several parallel pathways with a different number of synaptic delays are activated. This situation is common in the brain, especially in nonprimary sensory areas where the neural discharges are spread out temporally (for a comparison of temporal patterns in posterior auditory field and AI see Phillips and Orman 1984).

Compared with model 1, the maximum of \( \text{NTE} \) is only slightly lower for model 2 (Fig. 4, C and D). In contrast, the peak in the cross-correlation function, with three small peaks being each associated with one of the three delays used (Fig. 4E), has dramatically decreased from 0.6 to 0.2 (Fig. 4F). Interestingly, the maximum \( \text{NTE} \) occurs for \( \tau_{f} \) equal to the minimum delay (4 ms) and \( \tau_{p} \) equal to the maximum delay (10 ms). These properties of \( \text{NTE} \) emphasize its potential as a tool to investigate integration memory and information transfer in neural assemblies.

![Figure 1](https://example.com/figure1.png)

**Fig. 1.** Full causality between two Poisson processes: \( X_{2} \) is \( X_{1} \) delayed by 10 ms. \( \text{NTE} \) values, as a function of window size of forward integration \( \tau_{f} \) and past integration \( \tau_{p} \), are shown in A from \( X_{1} \) to \( X_{2} \) and in B from \( X_{2} \) to \( X_{1} \). C shows unbiased transfer entropy and D the preferred direction of flow (DF) from \( X_{1} \) to \( X_{2} \) as a function of \( \tau_{f} \) and \( \tau_{p} \).
The rationale for the need of shuffled estimates and normalization is emphasized in Fig. 5, where the quantities of transfer entropy are plotted for values \( \tau_f = \tau_p \) and \( \alpha = 0.5 \). The transfer entropy is increasing when \( \tau_f \) and \( \tau_p \) are increasing (Fig. 5A), as a consequence of the broad and sparse joint distributions. Removal of this bias ensures that the transfer entropy stays around 0 when no causality is present (Fig. 5B, dashed line for information transfer from \( X_2 \) to \( X_1 \)). Finally, because the amount of information available in \( X_1 \) and \( X_2 \) is also increasing (Fig. 5C), the normalization of the transfer entropy by this latter value sharpens the main peak at 10 ms. A higher average FR for \( X_1 \) or \( X_2 \) would basically have the same effect as increasing \( \tau_f \) and \( \tau_p \), i.e., increase of the amount of information available in \( X_1 \) or \( X_2 \). Consequently, the combination of bias removal, normalization (Eq. 4), and controlling the influence on the future of a channel of its own past (Eq. 1) makes NTE mostly independent of the firing rate of both neurons.

**Real data: information flow in cortical neural networks**

The ability of the transfer entropy to investigate neural assemblies is described in Fig. 6. Two arrays of eight electrodes are inserted in the auditory cortex of a normal hearing cat (Fig. 6A), array 1 being in a ventral part of AI where some recording sites showed nonprimary behavior (C3, C5, C6). This classification was based on longer response latency, more sustained responses, and nonmonotonicity, i.e., responses peak at an intermediate intensity level (Fig. 6B). For spontaneous firings, the matrix of NTE values (Fig. 6C) suggests various networks of information transfer graphically represented in Fig. 6D (for NTE >0.04). Little information was shared between electrodes in different arrays (Fig. 6C). A cluster analysis based on \( X_C \) values (Eggermont 2006) showed one cluster for array 2; one cluster consisting of C1, C2, C3, and C4; one cluster consisting of C5 and C6; and one consisting of a single electrode C7 (indicated by different colors in Fig. 6D). The maximum of NTE between electrodes from different arrays was consistently found for higher values of \( \tau_f \) and \( \tau_p \) (Fig. 6, E and F). Except for C10, a flow from left to right and bottom to top is visible in array 2 (Fig. 6D). Interestingly, a flow from primary to putative nonprimary recording sites is clearly visible for array 1 (Fig. 6D), and even associated with small \( \tau_f \) and \( \tau_p \) values (Fig. 6, E and F). The transitivity rule is respected.
here—that is, if there is no relation from channel 2 to 1, then there is none from 2 to 3, and from 3 to 1; however, some strong transfers might occur in both directions (for instance, C1 to C4 and C4 to C1, see Fig. 6C). One possible hypothesis is that this results from an indirect feedback. However, most information transfer occurs in one single preferred direction, as illustrated in Fig. 6G between C2 and C4 both with primary-like responses. For real data, just as for the simulations, a single peak is present in the surface of NTE values as a function of τ_\text{r} and τ_p (Fig. 6G).

One very interesting prospect for the transfer entropy is in the assessment of the neural assembly behavior for different auditory stimuli. A recent study demonstrated that correlated neural activity gives rise to clusters of neurons that expand and contract in size in response to different stimuli (Eggermont 2006). Such results may potentially be extended by means of an information transfer evaluation between arrays of neurons. The responses to different stimuli from the 16 recording sites described above strengthen this assumption (Fig. 7): the global network (i.e., the direction of flow and the recording sites involved) remains unchanged with regard to the stimulus used. However, there is a high variability across stimuli in the strength of information transfer between neurons of the network. For instance, there is more information transferred from site C13 to C7 when the stimulus is tonal and harmonic (Meow, RMeow, BaPa). In contrast, the maximum of information transferred from C16 to C10 is reached for clicks (Poisson, Train). In this particular case the cluster analysis resulted in only one cluster encompassing both arrays. An intriguing result also stems from the variability of τ_p across stimuli (Fig. 7B). Whereas close sites C9, C10, and C11 shared information for very small and unchanged τ_p values over all stimuli, distant sites C7 and C13 showed high variability for τ_p. More precisely, presentation of natural and altered Meows provoked more information transferred from C7 to C13 along with longer τ_p values. In contrast, pairs C9/C16 and C10/C16 also showed longer τ_p values during Meows or silence stimuli without any specific increase of NTE compared with other stimuli.

Real data: global results

Analysis of NTE values obtained for spontaneous activity from 5,650 electrode pairs in AI of 21 cats illustrates the putative statistical properties of the transfer entropy in vivo (Fig. 8). The distribution of NTE values approximately follows an exponential law (Fig. 8, A and B) with parameter \( \lambda = 1/0.0225 \), where 0.0225 is equal to the mean NTE. In particular, 5% of NTE values are \( > 0.0714 \) and 16% are \( > 0.04 \), the value taken as a lower limit in constructing the transfer diagram of Fig. 6D. The normalized information transfer computed without conditioning to the past of the current neuron was found to be 25% higher than NTE values, in average. This suggests that common history between pairs of neurons would account for roughly 20% of information transfer values if conditioning was not performed. The NTE values are somewhat correlated with peak cross-correlation values (Fig. 8C, correlation coefficient 0.58). Nevertheless, strong variability is apparent, suggesting the existence of pairs of neurons that are transferring information but are poorly synchronized, or in the opposite direction. This reflects the difference of information transfer revealed with these two tools. Similarly, the lag for the cross-correlation peak and the τ_p values are weakly correlated (0.31), although \( \tau_p \) is generally higher than the lag time in absolute value [\( P < 10^{-6} \), Wilcoxon test (Wilcoxon 1945); Fig. 8D]. This suggests that activity may be integrated over a larger interval than that strictly associated with the mean delay between neuronal firings.

Neural integration times

Figure 9 presents results about transmission times and neural integration times involved in information transfer in cat AI. Most of the influence of past activity was restricted to the next 5 ms of neuronal activity (distribution of τ_\text{r} values; Fig. 9A). In contrast, the duration of past integration memory was larger, generally extending \( F \) up to 15 ms but occasionally even \( F \) up to 35 ms (Fig. 9B). The highest values for NTE were found for integration-memory duration \( < 10 \) ms (Fig. 9C). As expected,
the information transfer decreased with distance between neurons (Fig. 9D), suggesting that, at least during spontaneous activity, redundancy between neuron activities occurs mostly locally. Consequently, the influence of multiunit activity from one recording site onto a distant one is weak and drowned in thousands of other incoming connections to this site. Another consequence is that the minimum \( NTE \) values increase with distance between electrodes (Fig. 9E). However, interestingly, some high values for \( NTE \) can be found even for nearby electrodes, suggesting the existence of neurons that process input activity over long temporal integration windows, even if in this case the \( NTE \) is necessary smaller. \( XC \) also decreases with distance between neurons in similar fashion to \( NTE \) (Fig. 9F).

**DISCUSSION**

**Causality and information transfer**

Most tools used in the investigation of causality in neuroscience, especially in electroencephalography (review in Gourévitch et al. 2006), are based on an interpretation of the Granger Causality definition (Granger 1969): “We say that \( X_1(t) \) is causing \( X_2(t) \) \( [X_1(t) \Rightarrow X_2(t)] \) if we are better able to predict \( X_2(t) \) using all available information than if the information apart from \( X_1(t) \) had been used.” In his paper, Granger interpreted “better able to predict” as a reduction of the variance of the prediction error. Yet, in the light of information theory, the ability to better predict can also be understood through the entropy of the predicted variable. If the uncertainty (entropy) associated with a random variable is reduced, the prediction of its possible values is indeed improved. From Eq. 1, it appears that transfer entropy is the reduction of uncertainty in the future of \( X_2 \) (\( X_2^T \)) attributed to the knowledge of the past of \( X_1 \) (\( X_1^T \)). Consequently, when common input does not explain all the activity, \( NTE \) is a quantification of a causality link in the Granger sense.

Because \( NTE \) is based on information theory, we also posit that it is a very general way to define causality, a way that encompasses both linear and nonlinear relationships between the activities of a pair of neurons. However, only bivariate cases are considered for \( NTE \) because “the information apart
from $X^P_1$ is $X^P_2$, and so all the available information is implicitly reduced to $X^P_1$ and $X^P_2$. We are aware that a “future challenge is to design methods that truly allow neuroscientists to perform multivariate analyses of multiple spike trains data” (Brown et al. 2004). However, even though NTE theoretically can easily be extended to an $n$-system of spike trains, it has been restrained in this paper to bivariate cases because of unobserved contributing neurons and the “curse of dimensionality” issues if all units available are used. One consequence is that “direct causality” should probably be avoided as an interpretation of NTE in a multiple spike-train context because of common inputs and potential intricate parallel and intermediate pathways between the pairs of neurons or multiple single units studied. A better interpretation may be that, in the Shannon sense, information present in one spike train is transferred by any synaptic pathway and subsequently observed in another train. Such a tool may thus be extremely useful in redundancy studies in the brain.

Neural assemblies

The greatest interest about neural networks in the brain concerns the parameters describing relations between neurons and their evolution during elicited responses. For instance, the

![FIG. 7. Variations of (A) NTE and (B) $\tau_p$ for several pairs of electrodes as a function of the sound used as a stimulus. All stimuli were used in previous studies (see METHODS).](image)

![FIG. 8. Statistical properties of NTE and comparison with peak cross-correlation (XC) on a database of spontaneous activity recorded with 5,650 pairs of electrodes in AI of 21 cats. A: distribution of NTE values. B: comparison of the density of NTE values with the theoretical model of an exponential law with parameter $\lambda = 1/0.0225$. C: scatterplot of NTE values as a function of XC values. D: $\tau_p$ values associated with maximum NTEs as a function of lag time at the peak of the cross-correlation function. One preferred direction was arbitrarily chosen for each pair of electrodes and values of $\tau_p$ or XC lag were thus of opposite sign when NTE and XC detected flows in opposite directions. This allows showing neurons pairs wherein a different direction of flow was found by the 2 methods. Only pairs of electrodes with NTE values $>0.04$ were used. A random value between $-1$ and 1 was added to both abscissa and ordinates for each point to help dissociate them graphically. Continuous line is the diagonal.)](image)
balance between inhibition and excitation appears crucial (Bush and Sejnowski 1996; Kirkland and Gerstein 1998; Xing and Gerstein 1996). It may drive the contraction or enlargement of neural assemblies observed through synchrony (Eggermont 2006). One of the most common definitions for neural assemblies is “a group of neurons [that are] at least transiently working together as indicated by correlation of unit activity” (Gerstein and Kirkland 2001). We feel that the restriction of assembly membership by correlation only is too limited. It seems to us that temporal integration—and thus information transfer as quantified by NTE—defines another parameter of relations between neurons that is also able to emphasize neural assembly properties. An extended definition of neural assemblies would rather become “a group of neurons that are at least transiently working together as indicated by significant levels of synchronization and short-time integration between their unit activities.” In this respect it is important to notice that the peak widths of the cross-correlograms (Eggermont 2000) are ranging over the same values as the integration times involved in NTE.

Presently, the size of microelectrode arrays (mostly 16 or 32 electrodes) does not allow exhaustive sampling of neural networks. However, it is likely that investigations in the next decade will produce hundreds of simultaneous recordings, from which more precise and realistic descriptions of neural assembly processing will arise. Regardless, NTE may be useful to make network models or neural computation models more realistic by defining additional physiological parameters (see, for instance, Bush and Sejnowski 1996; Davey et al. 2006; Feldman 1982; Graham and Willshaw 1997; Valiant 2006), especially those including temporal integration (Panchev and Wermter 2006) or feedback (Kirkland and Gerstein 1998; Xing and Gerstein 1996).

In particular, in the continuing debate opposing population codes based on firing rate with neural assembly code resulting from coincident spiking, NTE appears as a useful tool to investigate neural assemblies resulting from firing rate changes induced by temporal integration. Besides, results in Fig. 9E support the hypothesis of long temporal integration windows for some neurons even if small consecutive NTE values do not allow consistent conclusions. For extended studies of neural assemblies, it is likely that NTE can be used to complement the cross-correlation function.

Another important property of NTE dealing with neural assemblies concerns the conditioning to the past of \( X_t \) (\( X_t' \)), in the case of \( NTE_{X_t \rightarrow X_t} \). This conditioning cannot exclude common input that would provoke simultaneous activity in \( X_t' \) and

---

**FIG. 9.** Integration memory and influence of distance on information transfer. A: distribution of \( \tau_f \) values associated with forward memory (interval length influenced by the past activity). B: distribution of \( \tau_p \) values associated with past integration memory. C: scatterplot of \( \tau_p \) values as a function of NTE. Points are aligned on lines because of the discrete sample of \( \tau_p \) values used. D: NTE as a function of distance in millimeters between the electrodes. Log scale is used for ordinate axis. E: \( \tau_p \) values as a function of the distance in millimeters between the electrodes. Only pairs with NTE >0.04 were used. F: \( XC \) as a function of distance in millimeters between the electrodes. Log scale is used for ordinate axis. D, E, and F: continuous line shows locally weighted average.
X1p because of a delay between X1 and X2. Latency from the thalamus to a cortical cell is remarkably constant across the cortex (typically, ~2 ms), despite the wide divergence of inputs from the thalamus (Salami et al. 2003). This common input would thus occur without latency differences in cortical cell pairs. Somehow, it appears difficult to exclude it if the connection arising from common input overwhelms the strength of the direct connection between the pair of neurons. As previously noticed, it is one reason that NTE estimates should preferably be interpreted as an information transfer than as a direct causal link. However, the conditioning will exclude all common information between X1p and X2p. Not only is this important in the context of integration of activities between neurons, but because τ is most often greater than the lag detected by cross-correlation method (Fig. 8D), this will partly exclude the influence of similar values for X1p and X2p that would occur if the lag was small and X1p and X2p were determined by only a common input. We indeed found that 20% of information transfer between the only X1p and X2p arises from a common history between X1 and X2p and is removed by conditioning to X2p.

**Technical choices**

Applying information theory to any type of data always requires careful thinking about the parameters used. These parameters can indeed dramatically influence results and conclusions. We chose to directly use Eq. 2 and the data available to estimate the transfer entropy, keeping a nonparametric environment. Some closed forms for TE may exist albeit dependent on the model considered for data. For instance, using notations introduced in [METHODS], if X1 and X2 are Poisson processes, [X1(τ1)p]o, [X2(τ2)p]o, [X1(τ1)p]o, and [X2(τ2)p]o all follow a Poisson distribution. The computation of TE does thus depends only on the model used for the coupling relations between these four random variables. However, to our knowledge, such models have never been seriously considered in the literature and are even suggested as a future challenge in information theory context (Brown et al. 2004). As a consequence, the theoretical distribution of TE appears unreachable at this time, similarly to several causality measures recently proposed in electrophysiology [directed coherence or DCOH (Saito and Harashima 1981); directed transfer function or DTF (Kaminski 1999); partial directed coherence or PDC (Sameshima and Blinowska 1991); partial directed coherence or PDC (Sameshima 1999)].

A significance threshold for TE is also difficult to determine. One possibility is to use the work of Moddemeijer (1989), who basically noted that the histogram represents statistics following a multinomial distribution. He then proposed an approximation for the variance of the entropy estimate in the case of the histogram approximation of the density. Preliminary investigations adapting this idea to the TE statistic did not convince us of the robustness of such an approach, which too often gave significant values. We rather chose to normalize TE because a coefficient between 0 and 1 is easier to interpret, like that for correlation or coherence. In our case, NTE estimates the part of information conveyed by a channel that is independent of its own past but could be found in the past of another channel. The single TE statistic is indeed not comparable between channels because information conveyed by channels shows a high variability. We then computed the empirical distribution of NTE for spontaneous activity, which may be specific for the cat’s auditory cortex. Nonetheless, we think that values >0.03 or 0.04 could indicate real information transfer, albeit a modest one. The putative exponential distribution model for NTE (Fig. 7B) should help to delineate threshold values in future studies showing different NTE averages.

Another choice is the use of the same value τp for both the own past of a channel and the past of the exogenous channel. Obviously, it would be preferable to dissociate them, but the computation cost of an additional parameter to current τ and τp on which to maximize the NTE would be extremely high. One must notice here that this statistic in its current state already requires careful programming to achieve results in a reasonable time. In fact, the computation speed essentially depends on the joint distribution computation and thus on the number of trials used to compute the shuffled estimate TE(shuffled).

It is noted that similarities with the transfer entropy idea of conditioning with respect to the past of another spike train can already be found in the old “cross-intensity functions” (Cox and Lewis 1966; Perkel et al. 1967), although rarely used with neural data (Brown et al. 2004; Eggermont and Smith 1996), and in the nonlinear causality test of Baek and Brock (1992) improved by Hiemstra and Jones (1994). The mutual information between the synaptic input and the output spike train of a single neuron also was investigated by London et al. (2002) using a finite-order Markov model for sequences of activations. Transitional probabilities were estimated by means of a context-weighting tree representation of all possible models. Although complex, this method might represent an alternative to ours for entropy estimation, even if its ability to test several orders of memory and to manage a high number of pairwise combinations in a reasonable time remains to be proved.

**Physiological correlates of the results for spontaneous activity**

It is not surprising that transfer information is relatively low between cortical neurons (most values are <0.15; Fig. 8A), somewhat similar to maximum levels of synchrony under spontaneous activity (Eggermont 1994). Several histological reasons provide evidence for the weak influence of one neuron on another one, even neighboring ones. Even if one neuron typically receives inputs from several thousands of other neurons [rough estimates of 7,800 for mouse (Braitenberg and Schütz 1998), 9,400 for pyramidal neurons in rat visual cortex (Hellwig 2000), and 24,000–80,000 for human cortex (Abeles 1991)], it is much smaller than the total number of neurons in the brain [1.6 × 107 for mouse (Braitenberg and Schütz 1998), 1010 for humans (Abeles 1991)], even compared with the number of neurons that would be contained in the volume of the functional area of this neuron [75,000/mm3 in rat visual cortex (Hellwig 2000)]. Moreover, based on excitatory postsynaptic potential values in rat visual cortex (Song et al. 2005), around 26 presynaptic neurons would be needed to cause a postsynaptic action potential, a result that is within the estimate of between five and 300 (Abeles 1991).

Similarly, a decrease of transfer entropy and a fortiori synchrony with distance (Fig. 9, D and F) is consistent with anatomical findings. Hellwig estimated that 70% of synapses of layer 2/3 pyramidal neurons in rat visual cortex are con-
tained in a cylinder-shaped volume of cortex, whose radius parallel to the cortical surface is 500 μm and height is 300 μm (Hellwig 2000). Other studies led to similar results (Gruner et al. 1974; Nicoll and Blakemore 1993). Histological studies of Liley and Wright (1994) and Hellwig (2000) also showed decreasing connection probability with cell separation within pyramidal and stellate neurons of layer 2/3, the probability being <0.2 when the distance is >500 μm. The estimated probability of connection is often even lower in electrophysiology studies, between 5 and 15% for neighbor neurons (Mason et al. 1991; Nicoll and Blakemore 1993; Thomson and Deuchars 1997). Given that the mean synaptic delay in cortex is 1.2 ms with a minimum of 0.5 ms (Mason et al. 1991; Nicoll and Blakemore 1990), it also appears clear that large values for \( \tau_p \) (>10 ms) between two multiple single-unit recordings will be associated with distant connections and several synaptic intermediates. This will weaken the influence of the connection and the likelihood of similar activities, and so make NTE values decrease substantially (Fig. 9, C and D). Even if NTE and XC show a similar decrease with distance (Fig. 9, D and F), the variability observed between their values suggests that coincident spiking does not fully reveal the information transferred between neurons (Fig. 8C) and emphasizes the importance of part of the neural code based on temporal integration.

\( \tau_p \) values reported in Fig. 9 already provide an insight in the windows of temporal integration potentially used in AI. Figure 9E shows that long windows (>20 ms) can be found even between neighboring sites (<1.5 mm). Nevertheless, >80% of such observed windows are <15 ms. To our knowledge, most studies about potential temporal integration in auditory processing analyzed the responses to more or less complex stimuli, never under silence. For instance, some neurons in AI respond to brief periodic stimuli only for repetition rates <20–40 Hz (Eggermont 2002; Lu et al. 2001; Schreiner et al. 1997). Time reversal of short (<50 ms) segments in recorded speech does not affect its intelligibility (Saberi and Perrott 1999). The mutual information between some vocalizations and the neural firings in the ferret reached a maximum when the temporal resolution of the analysis was between 10 and 40 ms (Schnupp et al. 2006). From awake marmoset monkey responses to periodic click trains, Wang et al. (2003) concluded (Schnupp et al. 2006). From awake marmoset monkey responses to periodic click trains, Wang et al. (2003) concluded that rapidly modulated sounds would be integrated within a short-time window of about 20–30 ms. These observations suggest that temporal integration over 10 to 50 ms may occur when processing a more or less complex sound. These findings are thus completely in line when \( \tau_p \) values mainly between 2 and 15 ms found during silence, their distribution stretching up to 40 ms (Fig. 9C). One underlying question concerns the variation of NTE and temporal integration windows under various stimulus conditions. Figure 7, A and B showed that more information may be transferred between some recording sites during specific stimuli such as Poisson and Meows, whereas the length of the window of temporal integration is not perfectly correlated with variations of NTE. In particular, the study of spontaneous activity may be of more interest than previously expected because some significant levels of information transfer, and so redundancy, can be found between several multiple single units (Figs. 7A and 8A), even when they are >1 mm apart (Fig. 9C). This preliminary result is intriguing and illustrates the potential of the method in understanding certain aspects of brain processing.

In conclusion, normalized transfer entropy or NTE has promising features that should make it useful for neural networks analysis. Based on information theory and an intuitive definition, NTE quantifies the influence in a nonrestricted sense that activity observed in one neuron, or multiple single units, has on another one. NTE has great potential interest for studies of temporal integration as part of the neural code. NTE is a coefficient between 0 and 1 that is easy to interpret and independent of firing rate. NTE may show variability under various stimuli conditions, allowing studies of neural assembly encoding of stimuli. NTE appears robust (one peak over \( \tau_p \) and \( \tau_x \)) and shows results complementary to cross-correlation. NTE allows studies of feedback in neural circuits. Obviously, further investigations on NTE, \( \tau_p \) and \( \tau_x \) values between different places of a sensory cortical area, and during stimulus presentation may be needed to reveal the potential of this tool and possibly help to understand brain processing. The present application showed that most temporal integration windows during spontaneous activity in cat’s primary auditory cortex would extend from a few milliseconds to 15 ms.

**References**


**Grants**

This work was supported by the Alberta Heritage Foundation for Medical Research, the National Sciences and Engineering Research Council, a Canadian Institutes of Health–New Emerging Team grant, and the Campbell McLaurin Chair for Hearing Deficiencies.


