tripleC i(i): pp-pp, year
ISSN 1726-670X
http://www.triple-c.at

triple C cognition communication co-operation

Extraction of information of audio-visual contents

[a] *Carlos Aguilar. Universitat de Barcelona (UB). Facultat de Formació del Professorat. Campus Mundet. Edifici Llevant, 1ª P.; Passeig de la Vall d'Hebron, 171. 08035 Barcelona. Carlos.Aguilar@escac.es; http://www.ub.es/dev. Tel: 0034937361555*

[b] *Lidya Sánchez. Universitat de Barcelona (UB). Facultat de Formació del Professorat. Campus Mundet. Edifici Llevant, 1ª P.; Passeig de la Vall d'Hebron, 171. 08035 Barcelona. lsanchezg@ub.edu; http://www.ub.es/devp. Tel:0034934035151*

[c] *Manuel Campos. Universitat de Barcelona (UB).. Facultat de Filosofia. Departament de Lògica, Història i Filosofia de la Ciencia. C/ Montalegre, 6. 08001 Barcelona. mcamposh@ub.edu. Tel:0034934037992*

Abstract: In this article we show how it is possible to use Channel Theory [Barwise and Seligman, 1997] for modeling the process of information extraction realized by audiences of audio-visual contents. To do this, we rely on the concepts proposed by Channel Theory and, especially, its treatment of representational systems. We then show how the information an agent is capable of extracting from a content depends on the number of channels he is able to establish between the content and the set of classifications he is able to discriminate. The agent can endeavor the extraction of information through these channels from the totality of content; however, we discuss the advantages of extracting from its constituents in order to obtain a greater number of informational items that represent it. After showing how the extraction process is endeavored for each channel, we propose a method of representation of all the informative values an agent can obtain from a content using a matrix constituted by the channels the agent is able to establish on the content (source classifications), and the ones he can understand as individual (destination classifications). We finally show how this representation allows reflecting the evolution of the informative items through the evolution of audio-visual content.

Keywords: Information, Channel Theory, Audio-visual content, Situation Theory, Information representation

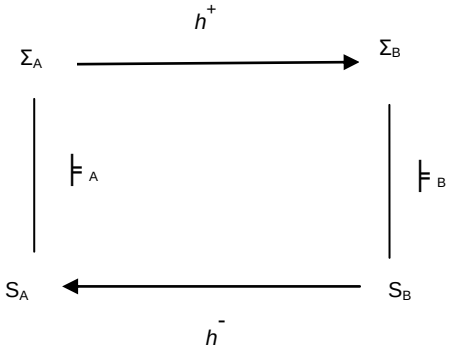
When an observer looks at the image of a house in the frame of a film, he can apply, for instance, a classificatory rule that allows him to separate the parts of the image that have the same color characteristics (we will not dwell on subtleties concerning color definition for frames, or thresholds). In this way, the observer obtains a series of elements that will correspond to objects identified as doors, windows, walls, etc. Moreover, these objects will appear throughout the film in a variety of states (orientations, sizes ...). In order to be able to state that the objects not only satisfy certain conditions, but also represent a certain sort of object, every observer will establish relationships between the extracted objects in every situation and some classification, such as a catalog of doors or windows. In this case, the visual interaction of an agent with the content and the catalog will enable the observer to establish identifications such as: object α in the frame turns out to be of the type of object β in the catalog. Channel Theory, however, does not identify the physical fact of vision with this channel. The concept of channel requires that a series of relationships governed by a local logic can be established for source and destination object pair, but doesn't specify which should be the physical basis of these relationships.

If we check how the flow of information is established in a system where we can set two classifications $A_1=\langle S_1, \Sigma_1, \models_1 \rangle$ and $A_2= \langle S_2, \Sigma_2, \models_2 \rangle$ we obtain as a result transfers of information with the form:

"$a_1 \models_1 \alpha_1$ contributes the information $a_2 \models_2 \alpha_2$".

Just as every ordered pair $a \models \alpha$ is relative to a classification, any transfer of information, which, in the end, is a 4-tuple $\langle a_1, \alpha_1, a_2, \alpha_2 \rangle$, is relative to an information channel. The theory posits the existence of an intermediate classification $B = \langle S_B, \Sigma_B, \models_B \rangle$ between classifications A_1 and A_2, which represents the set of physical interactions and abstract laws underlying the flow of information.

In order to link any part A of the system (from among A_1, A_2, A_3 ...) with kernel B, a couple of applications $h = \langle h^+, h^- \rangle$, with $h^+: \Sigma_A \rightarrow \Sigma_B$ in a direction and h^-: $S_B \rightarrow S_A$ in the reverse direction, are required, such that they satisfy $h^-(b)_A \models \alpha$ if and only if $h^+(b) \models_B (\alpha)$, for all $b \in S_B$ and for all $\alpha \in \Sigma_A$. We call this pair of applications *infomorfism* from A to B (and write: h: A → B). They represent a particular part-whole relationship in terms of information.



When we have two parts A_1 and A_2 related to the common core B through h_1: $A_1 \rightarrow B$ and h_2: $A_2 \rightarrow B$, we say that a situation $b \in S_B$ connects two situations $a_1 \in S_1$ and $a_2 \in S_2$ if and only if $h_1(b) = a_1$ and $h_2(b) = a_2$. Through this connection we manage to establish a relationship between concrete situations a_1 and a_2. In order to establish now an informative relationship between the types of situations α_1 and α_2 we need to consider the logical properties of A_1, A_2 and B. This process implies attributing each piece of reality its own logic, as elementary as it might be, as when we speak of the logic of a family structure or a thermal system. Formally we would say that, given a classification, we are interested in deductive relations between its types; relations that will depend on how these types classify situations. [Barwaise and Seligman, 1997] [Devlin 1991] [Pérez, 2007]

Given all this, we can say that an informative channel (or a channel of information) consists of an indexed family *C* of infomorfisms with a common condominium B, called *core of the channel*

$$C = \{h_i: A_i \rightleftarrows B\}_{i \in I}$$

With this definition, Channel Theory provides a mathematical model of the information flow that reflects the way in which agents reason about the world using partial information. This is, in comparison, much more than what Dreske's theory could do. On the other hand, both Shannon's mathematical theory of information as well as situation theory have been developed to be used in practical fields (signal transmission and computer science). Removing the machinery developed for dealing with types and restrictions, Channel Theory can be developed as an elegant mathematical theory. But, of course, it is precisely types and restrictions that are necessary to analyze the instances of information flow that happen in the real world. Therefore, the Channel Theory and situation theories constitute excellent approaches to the study of the flow of information. [Devlin, 2001]

1. Audiovisual content as distributed systems

We have introduced the notion of information channel

$$C = \{h_i: A_i \rightleftarrows B\}_{i \in I}$$

as a mathematical entity that gives shape to the flow of information in a distributed system. A whole made from parts in such a way that it allows for information to flow. In this scheme, the channel's core B represents this whole, A_i represents the classification of the parts, and the infomorfisms $A_i \rightleftarrows B$ represent the relation between the whole and the parts

When applying this representational scheme to a specific situation, such as the interpretation that a human agent makes of the "reading" of an audiovisual content, we can expect the appearance of a large number of classifications, as well as relationships, interacting in the information flow. The key question is how an agent can interpret this multiplicity of classifications and integrate them into a structure that can adequately represent, for the agent, the information stored in the content.

1.1. Information channels in an audio-visual content

Having established the general definition of the channels, we have to see how we can extend it to the flow of information holding between an audiovisual content and the agent who is viewing it (the term "viewing" does not refer to the physical process realized by the human eye on the content, but to the set of actions the agent realizes to capture what is delivered by the observed audiovisual content). In order to proceed to the establishment of this relationship, we must accept that the extraction of information from content by an agent can be modeled as a distributed system; more concretely, in a manner similar to the one through which the information is obtained by an agent when a representational system intervenes.

1.2. Representational system

If we restrict the definition of channel to the case of representational systems, we can see how these are a particular variety of channels. [Barwise and Seligman, 1997]

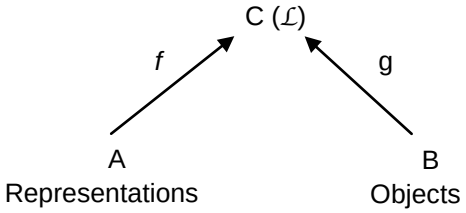
A representation system R = <*C*, $\mathcal{L}$ > is a binary channel

$$\mathbf{C} = \{\mathrm{f}: \mathrm{A} \rightleftarrows C\ ; \mathrm{g}: \mathrm{B} \leftrightarrows C\}$$

in which one of the classifications is designated as Source (A) and the other as Destination (B) together with a local logic $\mathcal{L}$, in the core of the channel *C*.

The representations of R are the objects of classification A. Let a ∈ tok (A) and b ∈tok (B); we say that object a is a representation of b ($a \rightsquigarrow_R b$), if a and b are connected by some *c* ∈ *C*. The object a is a faithful representation of b if a and b are connected by a normal object of *C*, some $c \in \mathcal{N}_{\mathcal{L}}$.

A set of types Γ, of the source classification, indicates a type β, of the destination classification ($\Gamma \Rightarrow_R \beta$), if the translation of the types in the channel core is a restriction in its logic $\mathcal{L}$, that is, $f(\Gamma) \vdash_{\mathcal{L}} g(\beta)$. The content of an object a, is the set of all the *types* indicated by its type of type. The representation a represents b as an element of type β if *a represents b* and β is in the content of a.



The connections between representations and objects shape the particular space-time process by which the representation is to represent what each object actually is (in the world of the agent who is interpreting it). The restrictions imposed by the logic model the different kinds of restrictions that may be introduced in the model. Conventional systems of representation, such as writing or images, involve a wide variety of different types of restrictions, ranging from the conventional to the physical. Audio-visual content as a representational system will establish restrictions in accordance with its nature; restrictions which will allow an agent to obtain information on the basis of the possibilities that the logic of these restrictions will allow him to obtain.

1.3. Establishment of channels on an audio-visual content

After establishing the foundation for the definition of channels and, particularly, of representational systems seen in the previous section, we show how we can establish, on an audio-visual content, a finite number of binary channels representative of the audio-visual content to study. The basis of all further information to be extracted from the content will come, basically, from the channels an agent may establish, from the result of applying these channels to the content, from the compositions of channels that can obtain among these channels, and from their temporal evolution of them.

A point distinguished from the application of the Channel Theory for the study of the information that an agent can extract from an audio-visual content concerns how it connects with the frameworks established by the systems of information retrieval and image recognition. It is sufficient to check the methods used by the systems of pattern recognition or the methods based on algorithms K-NN [Cover and Hart, 1967] orientated to the extraction of information from the images (static or in movement) to realize that the functional scheme fits with the theoretical approach of the informative channels.

The information to be extracted from a particular channel will be determined by the local logic (or, in general, by the set of local logics) that apply to it. This may result in the obtaining of different information for the same contents depending on the situations. This will happen if the agent can establish a channel between some of the source classifications of the content (Σ_i) and some of the destination classification (Γ_j) across a local logic supported by the situation. Again, we can see the parallelism with the image recognition systems; in particular, the recognition of the low-ranking characteristics of the content and the objects that it contains by means of strategies of categorization.[Biederman, 1987]

While the number of channels that could be obtained by following this reasoning is extremely high, of the order of n^m, it is undeniable that the number of channels that binary informative relations allow is finite and measurable. Going back to the content, it is possible for the agent (viewer) to consider as trivial some of the channels that can be established, since they are part of the ordinary process of meaning construction he performs. Let's see an example. We can establish, on a frame (still image), an information channel whose source types are connected regions of space sharing equal intensity, and destination types are the words that represent written English. If one browses through a channel thus set the frame with James Stewart's leg in plaster in the first scene of "Rear Window" (Alfred Hitchcock, 1954) we can obtain textual information established through this channel. "Here lie the broken bones of L.B. Jefferies".

Any information the agent may obtain from the same source classification (the connected regions of equal intensity) will always be based on the channels he may be able to establish (average intensity, position, orientation, ...). Similarly, we may either establish a channel, on some other frame of the same film, with the set of known objects such as cameras, chairs, cars, pictures, etc., or establish a unique channel referred to the types set as Objects. In short, the established channels will enable us to extract successive informative items; the independence of the information that each provides will have to be established a posteriori.

Once the processing of the still images that make up the content finished, we also have to consider how the sequencing of the images and the relation among the items obtained allows us to establish new channels. These channels do not work against the representations of a single frame, but against the development established by a set of them, as they appear in the order and cadence determined in the content. The establishment of these channels, of a more complex nature than binary representation systems, will be established from the variation and the relation of the items obtained through binary channels, and will require the development of new classifications Λ on these variations and /or relations.

The understanding of an audio-visual work, the extraction of its informational content, requires the definition of the space the agent has to model in his representation. We believe that this space transcends the standard definition of the concept of rectangular frame that applies to the window an agent can observe when he looks at the representation of the audio-visual product through a playback system. The cognitive experience, and, therefore, the element to be considered, should not be restricted to the 720x576 pixels of the player or the movie screen. One must consider, at least, the three-dimensional space associated with the situation depicted in the work at a given time (with one frame),

and must include all items that the agent is able to identify in this three-dimensional cube, including, of course, the experience of sound [Caballero, 2009]. Diverse systems of image recognition orientated to the extraction of information from the three-dimensional reconstruction of the space gathered in the image have been developed along the same lines. [Sminchisescu, 2006]

To model the common usage of language, and sound in general, through Channel Theory, we must also understand how it applies to the extraction of information from the sound part of the content. Let's consider how we can apply the process of establishment of an information channel in the case of speech. In a situation in which a speaker produces a sound, we can include that sound as an element of a classification; the classification established by the set of phonemes that exist in the language in which we are interpreting it. We can show that this classification is an infomorfism. Similarly, an infomorfism can be established between the graphic representation of phonemes in a language, and the set of phonemes of that language Therefore, both infomorfisms trivially establish a relationship with the common core of a classification: the one established by the language at stake (and the local logic that governs it), and we can therefore affirm the existence of an informative channel which, given a certain phonic sound, allows us to interpret it as referring to a particular sound, phoneme or graphical representation.

2. Extraction and representation of information

Once, the basis of reasoning that allows us to approach the study of audio-visual content using the formalism provided by Channel Theory established, we must determine which informative items we can extract from a content using Channel Theory, and how we can represent these.

Let's start by considering the nature of the object of study, audio-visual content. For an agent to establish informative channels for an audiovisual content he should be able to establish a series of infomorfisms between classifications. Therefore, we must start formulating the possible classifications we can establish on a content. The study of content as a whole would give us only information about the relations that can be established between the "object content" as such, and types that refer to content as a whole, such as classifications based on "gender", "format", "date" to mention the most obvious. It is clear, however, that the intrinsic information contained in an audiovisual work is distributed through its parts, through the relations established between them and the way they evolve. In order to obtain all this mass of information, it will be necessary to find which channels can be established between the components of the work and different classifications, and later to consider which other channels can be established given the evolution and the relation of its parts.

But how many channels can be established? We have seen in previous sections how, given *i* classifications, a channel can be established provided that an indexed family of infomorfisms between objects of the classifications and some common condominium B, called *the core of the channel*, can be established

$$C = \{h_i : \mathrm{A}_i \rightleftarrows \mathrm{B}\}_{i \in I}$$

Focusing on the case of channels that are systems of representation, for each source classification, we can establish a maximum of *j* channels on the possible object classifications as long as it be possible to establish the infomorfisms between the source classification Σ_i and the destination classification Γ_j.

Depending on the type of agent and on the situation in which the audio-visual content is included, the obtained values will be directly *information* (beyond the one established by the Theory of Channels). Let's think, for example, of a closed circuit of vigilance of an object, or a thermal camera that detects the temperature of the travelers in the airport. In both cases the final aim of these systems is simply the detection of low semantic level characteristics inside the video. Nevertheless, the systems orientated to the detection of concepts, actions or, in general, elements of high semantic level (as a human observer

who looks at a video) will have to be based on all the possible items (of high and/or low semantic level) directly extracted from the content on their attempt of managing to transcend the semantic gap.[Hauptman, 2007]

More generally, it would be possible to establish (if only in theory), for each source and destination classifications, some sort of channel. Exploiting this reasoning, we can express synthetically the set of channels we can establish using the expression:

$$\mathrm{C_{ij}} = \left\{\mathrm{h_{ij}}_k : \mathrm{A}_k \overrightarrow{\leftarrow} \mathrm{B}\right\}_{k \in K}$$

$$\mathrm{C_{ij}} = \left\{\mathrm{h_{ij}}_{k_1} : \mathrm{A}_{k_1} \overrightarrow{\leftarrow} \mathrm{B}; \mathrm{h_{ij}}_{k_2} : \mathrm{A}_{k_2} \overrightarrow{\leftarrow} \mathrm{B}\right\}$$

Therefore, seen in perspective, the number of representative channels an agent may establish can be represented by a matrix T_{ij}, where each row represents the set of channels that have been established for the *i*-th source classification (on the content), and each column will represent the *j*-th channel established for a given source classification Σ_i and a destination classification Γ_j.

$$T_{ij} = \begin{pmatrix} C_{11} & \cdots & C_{1j} \\ \vdots & \ddots & \vdots \\ C_{i1} & \ldots & C_{ij} \end{pmatrix}$$

When operating on the content we will obtain a matrix of informative values resulting from the information channels we have been able to establish on it. The application of this matrix on the audio-visual content will result in a matrix of values:

$$T_{ij}\,(\text{audio} - \text{visual content}) = V_{ij}$$

$$V_{ij} = \begin{pmatrix} Val_{11} & \cdots & Val_{1j} \\ \vdots & \ddots & \vdots \\ Val_{i1} & \ldots & Val_{ij} \end{pmatrix}$$

Each value Val_{ij} will, in fact, be a vector of the values information channel C_{ij} has returned when applied on the content. Seen from the opposite point of view, each object which, according to the source classification Σ_i, may establish an informative channel with the successive classifications Γ_j will provide an element of matrix of informative values. So if we think of each of the informative values obtained as an element of the matrix formed by the source and object classifications we will have a representation of the information supported by the content studied.

For any two agents who establish different representation systems, we will be able to establish a transformation between the matrices that, for each of them, represent the information, as soon as we are able to obtain the channels that are established between the classifications of both agents. In order to simplify, let's consider the case in which two agents establish the same sort of classifications on the content Σ, but are able to establish different object classifications Γ y Γ'. In this case, if we can determine the set of channels that can be established between Γ y Γ' (T'_{jk}), we will be able to find the equivalence between the two matrices that represent the informative items of the audio-visual content for each one of the agents.

$$T_{ik} = T_{ij} \cdot T'_{jk}$$

The elements of T_{ij} only combine with those of T_{jk} when a channel composition is established that allows for it in accordance with the Xerox principle [Barwaise y Seligman, 1997], yielding, otherwise, null values.

2.1 Evolution of information

The only informative relations an agent can get to know using the sort of reasoning described above are the informative relations obtained at the level of representation, but these are not the only possible ones. It will be possible to establish the existence of new channels of information from the establishment of new infomorfisms between elements obtained in the information matrix and between the classifications that can be established on this matrix and the corpus of classifications established by each agent. This

is the path on which we should build the semantic and syntactic knowledge of the content represented in the audio-visual work.

The representation of the constituent elements of the work (image/sound) as they evolve over time will be reflected by the variation of values (through time, or any other parameter) of the representation matrix. This representation and its evolution are supported by the situation in which the viewing of the work takes place, and will be valid only for it. The condition *sine qua non* for understanding the informative items obtained is its subordination to the conditions established by the situation. For instance, given a frame, if the situation in which it is inscribed varies, the channels that will be established will also vary, since the logic that governs the core of these channels is tied to the conditions of the situation.

In a first approximation we can assume that the channel matrix can be represented as stable over a time span. Depending on the situation this may be a frame, a shot, a scene or an arbitrary subdivision of time. Once this premise accepted, the informative values obtained from the establishment of information channels can be represented by a time function representing the evolution of these values.

$$V_{ij}(t) = \begin{pmatrix} Val_{11}(t) & \cdots & Val_{1j}(t) \\ \vdots & \ddots & \vdots \\ Val_{i1}(t) & \dots & Val_{ij}(t) \end{pmatrix}$$

The value matrix itself will also evolve through time, but in this case the channels may or may not exist, so that each element C_{ij} (t) can yield the null value for a certain instant if the condition for the existence of the channel doesn't hold. In short:

$$C_{ij}\ (\mathrm{t}) = \begin{Bmatrix} \nexists\ if\ \nexists\ h_{ij}\ \text{for the classfication A}_{ij} \\ \exists\ h_{ij}\ \text{for the classfication A}_{ij} \end{Bmatrix}$$

3. Conclusions and future development

We have shown how an agent can extract information from an audio-visual content following a theoretical formulation marked by Channel Theory. After showing it, we have established a possible method of representation of all the informative values an agent can obtain from a content using a matrix formed by the channels the agent is able to establish on the content (source classifications) and those that he, as an individual, is able to understand (destination classifications). Finally, we have shown how this representation allows us to show the evolution of informative items through the evolution of audio-visual content. The next objective of this research is to assess how the variations of the information matrix can allow the obtaining by the agent of new classifications and channels that go beyond the objects represented in the content, and can also allow an approach to the understanding of the semantic and syntactic levels of content by the agent.

References

Barwise, J., Seligman, J. (1997). *Information Flow: The Logic of Distributed Systems*. Cambridge: Cambridge University Press.

Biederman, I. (1987). Recognition-by-components: A theory of human image understanding. Psychological Review. Vol 94(2), Apr 1987,pp 115-147.

Caballero, J.J (2009). El entre como espacio generativo de la expresión fílmica. (Doctoral thesis -Universidad de Barcelona)

Cover, T. M., Hart, P. E. (1967). Nearest neighbor pattern classification. IEEE Transactions on information Theory, IT-13, 21-27.

Devlin, K. (1991). Logic and Information. Cambridge, UK: Cambridge University Press.

Devlin, K. (2001). *The Mathematics of Information*. [On line] Helsinki (Finland): European School of Logic, Language and Information. <http://www.helsinki.fi/esslli/courses/Logicinfo.html> [Consulta: 18/12/2009]

Hauptman, A. (2007) How many high-level concepts will fill the semantic gap in video retrieval? In proceedings of the ACMInternational Conference on Image and Video Retrieval, 2007.

Pérez-Montoro, Mario (2007). The Phenomenon of Information. A Conceptual Approach to Information Flow. Lanham (Maryland): Scarecrow Press. ISBN 978-0-8108-5942-5.

Sminchisescu, C. Triggs, B. (2001) Covariance Scaled Sampling for Monocular 3D Body Tracking. Proc. IEEE Conf. Computer Vision and Pattern Recognition, vol. I, pp. 447-454, 2001.

About the Author

Carlos Aguilar is current Ph.D Student for the DEVP department of the School of Education at the Universitat de Barcelona. Currently teaches in the ESCAC (Escola Superior de Cinema i Audiovisuals de Catalunya).

Lydia Sánchez is Ph.D. in Philosophy by Stanford University. She is currently teaching Communication in the DEVP department of the School of Education at the Universitat de Barcelona. She has taken part in the doctoral programs *Comunicació, Art, Educació* and *Formació del professorat: pràctica educativa i comunicació*. She also teaches in the Humanities, Psychology and Documentation departments of the UOC. Her research focuses in the theory and philosophy of communication. She is author of different articles and contributions to books, and co-edited “Industrias de la comunicación audiovisual” (2008).

Manuel Campos is Ph.D. in Philosophy by Stanford University. Currently teaching at the Logic Dept. of Universitat de Barcelona. Interested in Philosophy of Language and Communication, and Philosophy of Science.