**“VISUALISATION OF MUSIC” INSPIRATIONAL SESSION PROPOSAL**

# A learning strategy for audiovisual composition

Rodrigo F. Cádiz and Patricio de la Cuadra
Centro de Investigación en Tecnologías de Audio, Instituto de Música
Pontificia Universidad Católica de Chile
rcadiz@uc.cl, puadra@uc.cl

Recent technological developments have enabled us to synthesize images and sounds concurrently within single computers, even in real-time, giving birth to novel and genuinely integrated audio-visual art forms. But, how should we organize and compose such works? In other words, given a certain soundscape, what would be an appropriate sequence of images to that soundscape? Given a certain sequence of images, what soundscape is appropriate to it? If the image sequence and the soundscape are being created concurrently, how should we compose them?

Authors have proposed different approaches to these questions. These approaches differ significantly, and they are based on diverse principles, such as correspondence of aural to visual harmony, audio-visual modeling of mathematical principles, audio-visual rendering, data sonification, algorithmic control and parameter space exploration. It is important to notice that there is no easy or correct solution because the problem we have to deal with is that of combining two entirely different media in time, and whose relationships can be so complex that they cannot be accurately modeled or even comprehended.

We are interested in a mapping strategy inspired by two fundamental ideas: isomorphism and synaesthesia. Isomorphism applies when two complex structures can be mapped onto each other, based on the fact that changes in one modality consistently cause changes in another modality. The word "synaesthesia" comes directly from the Greek syn "together", and aísthesis "perceive", thus meaning "a union of the senses". Synaesthesia occurs when stimulation in one sensory modality automatically triggers a perception in a second modality, in the absence of any direct stimulation to this second modality.

We propose a learning-based approach to audiovisual composition. This approach is based on a neural network model that enables the learning of flexible mapping of either aural or visual information into the other, and it is able to generate complex audio-visual relationships by training. This approach allows the possibility of generating new audiovisual art forms based on past experiences. Given a set of audiovisual relationships, the proposed system can learn these relations and produce new ones when exposed to new unknown inputs.

This approach has been implemented in software in the widely known environment Max/MSP/Jitter and both the training and computation of the neural network can be done in real-time.