A fuzzy logic model for compositional approaches to audio-visual media

Rodrigo F. Cádiz
Music Technology Program, School of Music, Northwestern University
rcadiz@northwestern.edu

Abstract

A fuzzy logic approach to the challenge of composing both sound and moving image within a coherent framework is proposed. This approach is based on a fuzzy logic model that enables a flexible mapping of either aural or visual information into the other. This model is inspired by synaesthetic experiences, in which a stimulus in one sensory modality triggers a perception in a second modality. In the proposed model, both aural and visual information are parameterized and classified into several fuzzy sets. These sets are used as either inputs or outputs to a system that uses 26 fuzzy rules to control the behavior of the mapper. Once the rules are computed the variables are defuzzified and converted into the corresponding output parameters. “ID-FUSIONES” (2001), an audio-visual composition based on this model is presented and discussed as an actual implementation of this approach.

1 Introduction

Recent technological developments have enabled us to synthesize images and sounds concurrently within single computers, even in real-time, giving birth to new 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?

Authors have proposed different answers to these questions (see Kim and Lipscomb (2003), Lokki et al. (1998), Hunt et al. (1998) and Rudi (1998)). It is important to notice that there is no easy or correct answer because the problem we have to deal with is that of combining two entirely different media in time.

The approach presented here is inspired by structural isomorphisms, i.e., information preserving transformations. 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 (for a full discussion of these topics, please refer to Hofstadter (1999)).

It is important to mention that the proposed model does not work in real-time, although it can be modified to incorporate this capability in the future.

2 Experiencing the audio-visual

In order to understand how the composition of an audio-visual work should be addressed, it is important to take into consideration how presenting music with visuals affects the listener differently from music alone.

According to Bullerjahn and Guldenring (1994), Iwamiya (1994), Lipscomb and Kendall (1994), Rosar (1994) and Sirius and Clarke (1994) there is substantial empirical evidence that music and moving images interact in powerful and effective ways. However, as Finaas (2001) suggests, it is often hard to predict the exact influence of the visual stimuli that relates to audio stimuli. Is it well known that vision is a powerful sense and that visual stimuli may interfere in a negative way with listening to music. The visual component may not only fail to enhance the listener’s experience but it may attenuate the aural experience compared to that resulting from just an aural listening.

The visual elements and their relationship to the music can vary tremendously and studies such as the one described in Sirius and Clarke (1994) show that the effects of music in the rating of visual images are usually additive and that there are no interactions between specific musical styles and particular visual images.

In his article, Finaas (2001) suggests three different categories for audio-visual presentation of music: Simple documentary, which is just a simple exposure of the live performance, TV-type, where live performance is alternated with images from various perspectives, close-ups and images of details and Non documentary, which does not aim to be a
faithful description of the performance.

In the present work, I am concerned only with “non documentary” types of presentation, in which the moving images do not necessarily resemble the performance or the score of the music being presented. This kind of audiovisual interaction is inspired by the perceptual phenomenon known as synaesthesia.

3 Synaesthetic art

"Synaesthesia is defined as an involuntary physical experience of a cross-modal association. In other words, there is a crossing of the senses. Having one sense stimulated would cause a stimulation in another sense as well. This would be like a ringing bell being heard and seen. For those who possess synaesthesia it is an obvious and integral part of their sense perception. For a synaesthetic artist there is no question about using it or not, because it is simply there. But to those who do not possess synaesthesia it remains complex and very often misunderstood (Berman 1999).

Works that fuse the senses are often referred to as synaesthetic art (Hertz 1999). This is possibly due to the appropriation of the term synaesthesia by many art historians who have written about the interrelationships of music and art (Berman 1999).

However, it is very important to distinguish the term synaesthesia from the term synaesthetic art. As Paul Hertz suggests, “Synaesthetic art is a deliberate contrivance, the product of an artistic aspiration, and we should not confuse it with the neurological phenomenon of synesthesia. Persons, not artworks, are synesthetic” (Hertz 1999). Also, these kinds of artwork are usually created for an audience that is not synaesthetic.

Cytowic also states that there is a sharp demarcation of synaesthesia as a sensual perception, as distinct from a mental object like cross-modal associations in non-synaesthetes, metaphorical language or even artistic aspirations to sensory fusion (Cytowic 1997).

The fuzzy logic mapper described in this article does not necessarily constitute a methodology for creating synaesthetic art, but it is rather inspired by synaesthesia as a perceptual phenomenon. Perhaps it is more appropriate to think about it in terms of “cross-modal” art. Synaesthesia consists not of random associations between isolated phenomena or qualities of two sensory domains, but rather expresses correlated dimensions or attributes (Marks 1997). In this sense, synaesthesia acts like a mapping from one modal dimension to another, and this is the idea that inspired the work described herein.

4 Fundamentals of Fuzzy Logic

Everything is a matter of degree. This statement is known as the Fuzzy Principle (Kosko 1993) and it is one of the main issues in fuzzy logic theory. In 1965, Lofti Zadeh published a paper called Fuzzy Sets, in which he applied multivalued logic to sets or groups of objects. This famous paper gave birth to the field of fuzzy logic as we know it today.

Fuzzy logic systems have been widely used in engineering and control applications. Japan is the world wide leader in fuzzy products (Kosko 1997). The most famous Japanese fuzzy application is the control of the subway system in the city of Sendai. Fuzzy logic has been successfully used in products such as cameras, camcorders, washing machines, vacuums, microwave ovens, braking systems and even an unmanned helicopter. Despite its more than 40 years of applications in engineering, economics and social sciences, fuzzy logic has seldom been used in the artistic and creative fields.

4.1 Crisp logic versus fuzzy logic

In crisp logics, such as binary or boolean logic, variables are either true or false, black or white, 1 or 0. Aristotle’s law of the excluded middle holds: A or not-A (Kosko 1993). A given thing cannot be part of A and not-A at the same time. If so, it’s a contradiction. In contrast, fuzzy logic is defined with uncertain terms and partial values of truth are admitted. Things are not true or false or black and white anymore; they can be partially true or false or any shade of gray. Mathematically, fuzzy logic accepts values between 1 and 0.

4.2 Fuzzy sets

Theoretically, a fuzzy set \( F \) of a universe of discourse \( X = \{x\} \) is defined as a mapping, \( \mu_F(x) : X \rightarrow [0, \alpha] \) by which each \( x \) is assigned a number in the range \([0, \alpha]\). When \( \alpha = 1 \), which is the usual, the set is called normal. In the extreme case where the distribution is of zero width, the membership function is reduced to singularities, i.e., the fuzzy set reduces to a crisp set. If the singularities are of two possibilities, we have binary logic. \( \mu_F \) is the grade of membership (or degree of truth) of \( x \). The most commonly used mathematical functions representing membership curves are triangular and Gaussian functions.

4.3 Fuzzy rules

Fuzzy logic attempts to emulate the way humans reason with vague rules of thumb and common sense. Human beings make decisions based on rules. Even though we may not be aware of it, all the decisions we make are based on computer-like if-then statements. For example, if the weather is fine,
then we may decide to go out. If the forecast says the weather will be bad today, but fine tomorrow, then we make a decision not to go today, and postpone it until tomorrow.

Rules associate ideas and relate one event to another. Fuzzy rules also operate using a series of if-then statements. Fuzzy rules define fuzzy patches, a key concept in fuzzy logic. A machine or system could be made smarter using what Kosko (1993) called the Fuzzy Approximation Theorem, idea that is based in the use of several fuzzy patches as an approximation to any given mathematical function.

4.4 Fuzzy system

A fuzzy system is defined as a system with operating principles based on fuzzy information processing and decision making. In such a system both inputs and outputs are classified and de-classified into fuzzy sets. The process of classification is defined as fuzzyfication and the de-classification as defuzzyfication. Once the inputs are fuzzyfied several fuzzy rules are computed in parallel to produce the corresponding outputs. These outputs are then defuzzyfied so that the desired variables are obtained. A fuzzy system is capable of reproducing the same output for the same input parameters at any given time. Several methods and techniques for fuzzyfication and defuzzyfication are proposed in the literature, each one with its advantages and disadvantages.

4.5 Why go fuzzy?

Fuzzy systems are powerful and work in a way that resembles some characteristics of human behavior. Parallel computation of fuzzy rules reduces drastically the computation time compared to a traditional mathematical approach. Fuzzy systems allow approximation of highly non-linear systems with incredible accuracy. One of the nicest things about this is that it is not necessary to know any mathematical model in advance to approximate any system. Fuzzy logic allows us to build systems using common sense, and the fuzzy rules can be discussed, tuned, and detuned easily.

5 ID-FUSIONES, an audio-visual work based on a fuzzy mapper

ID-FUSIONES is an audiovisual work composed by the author and graphic designer Luz María Cury in the period May 2000-March 2001. This work is completely constructed on the basis of successive events in time and each one of these events has two instances, one in the aural space and another one in the visual. Each one of these instances has four associated variables or parameters. However, each event is unique and has eight variables that define it completely. The proposed model in this paper is able to translate these parameters from one space (visual or aural) into the other.

In this piece, the aural parameters were considered inputs and the visual outputs. This could vary with different applications of the fuzzy model. The relationship could be inverted (visual as input and aural as output) or even a mixed approach could be used (some visual and aural as inputs, some visual and aural as outputs).

5.1 Parameters

At this point it is important to mention that these parameters, only eight, were chosen for simplicity reasons. This might seem a contradiction, as fuzzy logic models are one of the better tools to avoid simplicity. In theory, the proposed model could take any number of input and output parameters with no limitation other than computational power. However, only eight parameters were used in order to make the compositional process simple. These parameters also coincide with those identified to be perceptually relevant in the literature. Seashore (1967) mentions that "...sound waves have four, and only four, characteristics; namely, frequency, amplitude, duration and form. Sounds of every conceivable sort, from pure tone to the roughest noise, can be recorded and described in terms of these four". By the term “form”, Seashore is referring to timbre. Lipscomb (1995) found pitch, loudness and timbre to be perceptually relevant in the aural domain; and location, shape and color in the visual domain.

The chosen parameters for the model are described below.

Aural parameters

- **Frequency** Corresponds to the actual main or fundamental frequency component of the spectrum of each sound event measured in Hertz.
- **Intensity** Corresponds to the initial intensity (attack) of each event measured in decibels (dB). In this work, levels between 60 and 100 dB were used, as a way to cover all the usual dynamic range.
- **Duration** Each event has a temporal duration that determines its existence. The duration of each event is measured in seconds. Although this parameter is considered an input to the fuzzy mapper and an aural variable, it also determines the duration of the events in the visual plane. This is arbitrary, because the fuzzy model is flexible enough to allow a not one-to-one correlation.
• **Noise** Sound events were considered *noisy* or *not noisy* according to the shape of the spectrum. The flatter the spectrum, the noisier the sound.

### Visual parameters

• **Color** Initially, the entire light spectrum was considered, red to violet. This range was subdivided into 16 colors plus two distinct shades of gray. This parameter is highly associated with pitch and loudness (Kim and Lipscomb 2003). But, once again, the flexibility of the fuzzy model allows to have a more complex relationship between color, frequency and intensity, so that the trivial relation, *higher the frequency, higher the color* is avoided.

• **Shape** Each event has a certain shape, classified on a perceptual complexity scale between 0 and 1.

• **Size** Each event has a certain size on the screen. This parameters could be small, medium or large.

• **Motion** All the visual events of the work are in constant movement. Motion, in this context, is equivalent to velocity, because it could have a certain direction and a certain speed.

### 5.2 Fuzzy sets

Once the eight parameters were defined, it was necessary to classify them into fuzzy sets or *membership functions*. The majority of the variables were classified into very low (VL), low (L), medium (M), high (H) and very high (VH) regions. Although some of the parameters were fuzzyfied taking into consideration some of their perceptual characteristics, it is important to emphasize that this process could have been totally arbitrary. The proposed fuzzy model is mainly a creative tool and it was not designed to reproduce perceptual or physical phenomena.

Figure 1 shows the membership functions for the aural parameter frequency. As it can be seen, frequency is classified into five different fuzzy sets. Triangular functions were used for each set. Note the uneven distribution of the size of the various sets, which resembles a logarithmic function. This was done to reflect the actual perception of *pitch* by the auditory system. Figure 2 shows the membership functions for intensity. As with frequency, triangular shapes were used and the uneven size distribution was done to reflect the perception of *loudness* (in dB).
Figure 5 shows the membership functions for color. Gaussian functions were used in order to emulate the actual frequency distribution of the visible spectrum. Figure 6 shows the membership functions for shape. Evenly spaced triangular functions were considered. Figure 7 shows the membership functions for size. Only three fuzzy sets were used in this case, classifying size into small (S), medium (M) and large (L). Figure 8 shows the membership functions for motion. Evenly spaced Gaussian curves were used.

5.3 Fuzzy rules

The relationships between the different variables were regulated by means of fuzzy rules of decision, that control the behavior of the whole mapper.

A rule of decision of this type could be for example:

If the frequency is very low then the color is very low, the shape is somewhat complex, then the size is big and the motion is medium.

or:

If the frequency is very low and the amplitude is high or the sound is noisy then the color is gray, the shape is somewhat simple, then the size is big and the motion does not really matter.

Figure 9 shows a graphical summary of the 26 fuzzy rules used in this work. The first four columns show the input (aural) variables and the last four the output (visual) variables.

Proceedings ICMC 2004
Inside each rectangle the membership function that is affected by each rule is shown. All these rules were computed in parallel and then defuzzified to create the actual output values for each visual parameter.

The parallel computation of these 26 fuzzy rules corresponds to an eight-dimensional non-linear function. However, the system could be represented by several three-dimensional functions that relate two inputs to a given output. Figures 10, 11 and 12 show some of these functions according to the 26 fuzzy rules used in this piece and discussed previously.

### 5.4 Screenshots

Figures 13, 14, 15 and 16, correspond to screenshots of the piece taken at different times. Additional screenshots and excerpts are available at [http://www.rodrigocadiz.com/idfusiones](http://www.rodrigocadiz.com/idfusiones).

**Figure 10:** Three dimensional representation of the relationship between frequency, intensity and color

**Figure 11:** Three dimensional representation of the relationship between frequency, duration and motion

**Figure 12:** Three dimensional representation of the relationship between intensity, noise and size

**Figure 13:** Screenshot at 2:10

**Figure 14:** Screenshot at 5:29
6 Limitations and future work

The presented implementation does not work in real time and it can only generate non-interactive mappings. A very important direction for future work in this area would be to extend this system to a real-time based framework. This would allow, for instance, generation of visual images on the fly in reaction to live music or vice-versa.

Another possible direction for future research would be to develop a neuro-fuzzy mapping model. Such a system could be able to learn how to react to a given unknown input, either aural or visual, and generate a mapping based on the knowledge acquired on past experiences.

7 Acknowledgments

Thanks to Luz María Cury for all her help and support. This work was partially funded by a grant for artistic creation and research of the Andes Foundation, Santiago, Chile. Thanks also to Dr. Gary Kendall and Dr. Scott Lipscomb for their valuable comments and suggestions.

References


