STRUCTURAL DEVELOPMENT FOR A COGNITIVE ARCHITECTURE IMPLEMENTED IN RELATIONSHIP OF AUDIO-VISUAL SYSTEM

Sebastián González, *IA889 - Sistemas de Cognição Artificial, Seminário Interno de Cognição Artificial - SICA 2014, teacher RICARDO GUDWIN.*

Abstract—Through the systems of artificial cognition, this project is a tool oriented to understanding the phenomenon, that links the representations of the sound world in visual images. A system that can be trained to establish development of rules for audiovisual association. the associations may be general and customized for the user.

Index Terms—audio, visual, audiovisual, psychoacoustics, visual perception, cognition, architecture, relationship, artificial, transformation, synesthesia

1 INTRODUCTION

Takete and Maluma [1] - Booba and Kiki: [3]

Figure 1

THE effect was discovered by Wolfgang Köhler in 1929, observed the human brain attaches abstract meanings to the shapes and sounds, 95% to 98% selected the curvy shape as "bouba" and the jagged one as "kiki" viewed in Figure 1; Daphne Maurer [2] shows that children two and a half years old may show this effect as well.

If the possibility of creating a relationship between the sound of a word and an image is true, it is possible See the Sound?

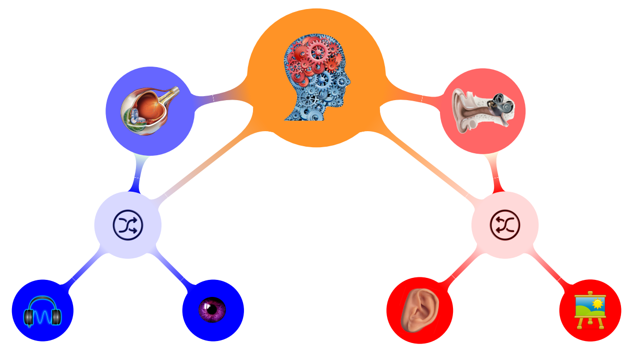

Figure 2: Audio-Visual Relations

This proposal is the beginning of a project to develop a computer system, the purpose is to relate the sounds and images using artificial cognition architectures. A artificial cognitive systems is implemented to serve as an intermediary to contribute in the construction of auido-visual relationship with the user.

The system input is an audio signal, then it is processed to create audio-visual relationship to generate animations in the output. The goal is to establish audio-visual relationships that they can be used generally for people, but also get personalized relationships for each user.

Also the extent that can be is the objective is to create a transformation for audio-visual correlation to establish a communication between hearing and non-hearing people.

- *IA889 - Sistemas de Cognição Artificial, Seminário Interno de Cognição Artificial - SICA 2014, teacher RICARDO GUDWIN.*
- *Sebastián González, E-mail: mail segodimo@gmail.com*

UNICAMP – Programa de Pós-Graduação FEEC - 2014

2 DEVELOPMENT PROPOSAL

2.1 QUESTIONS AND PROBLEMS

To reveal the development of the system is necessary to categorize and solve the following questions nd problems:

- How to view or Sound?
  - What relationships must exist to See or Sound?
    - Define Input Parameters in data form.
    - Define elements for perception.
    - Define elements for attention.
    - There representation of information? (Semiotics).
  - Define output parameters in data form (Output Images Type).
- Establishment of rules relating to convert Sound to Images.
  - Relationship levels of the senses.
  - Building relationships das levels.
  - Creating the transformation of ratio and levels.
    - First level: Psychoacoustic Vs. Visual Perception.
    - Second level: Interpretation of emotion and motivation.
    - Tertiary: Semiotic Icon - Index - Symbol.
- Define the types of memory used in the system audio visual relationship?
  - Sensory memory
  - Working memory
  - Long term memory
- Learning
  - Adapt the system to the user (Adapted Relations with Machine Learning).
  - Pattern Recognition.
  - Meta-cognitive learning.
  - Evaluated Learnings Acquired.
- What the Goals of the system?
  - What are the Results of the system?
  - What Elements Output?
- What will be the Behavior of the system?
  - Actions
  - Tasks
  - Plans
  - Purposes
  - Motivations
  - Emotions
- Definition of cognitive architecture for the system?
- Frameworck definition to use.
  - LIDA framework
  - SOAR framework

2.2 COGNITIVE SYSTEM MODULES

In the construction of the cognitive modules, it is proposed to start each adapting it to the focus of audiovisual transformation. tackled in this article are the first modules of attention perception and categorization. It's wants to study each of the modules in a cognitive architecture progressively between planning and implementation, so that from the possible questions that arise in the development can generate the following construction concepts.

Then a diagram of the modules that can be addressed.

Figure 3: Cognitive Modules

2.3 AUTONOMOUS AGENTS AS EMBODIED AI (FRANKLIN 1997)

Franklin's diagram is used as reference point to identify the modules to be used in the development of cognitive architecture for audiovisual relationship. Then are associated with the cognitive modules (*Figure 3*) that must be addressed.

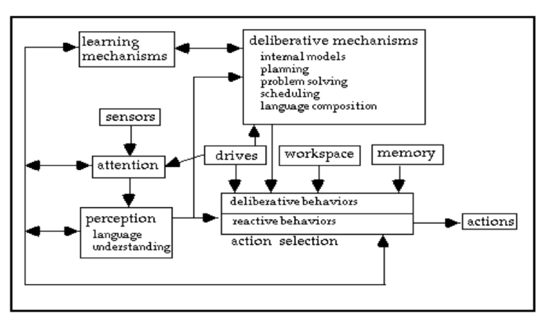

Figure 4

This is a visual interpretation of Franklin, the reason for this diagram is to display the cognitive modules (*Figure 3*) in the Franklin's architecture.

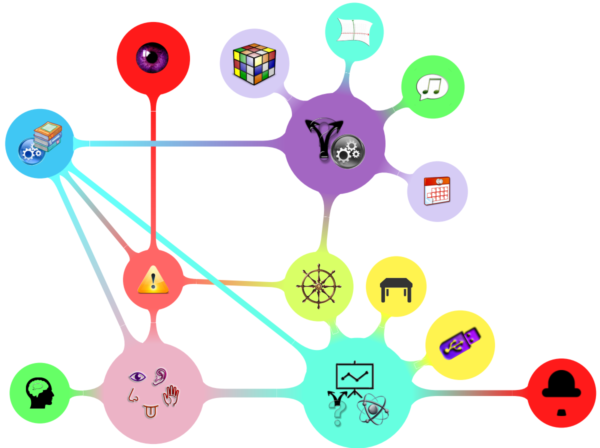

Figure 5

2.4 COGNITIVE ARCHITECTURES

This is the first model to create cognitive architecture:

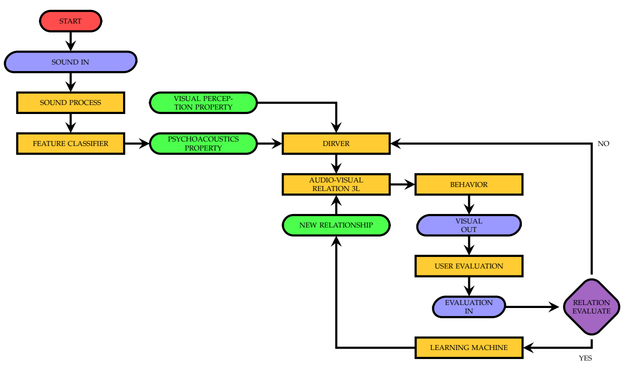

Figure 6

The first starts with an audio signal input, then the signal is processed and classified in psychoacoustics events such as data, these data will be stored in a memory field of architecture; These are the basic parameters of sound categorization to create the relationship.

There exist fixed parameters and relative parameters. Fixed parameters belong to the laws of Auditory (*psychoacoustics*) or Visual (*perception*). The relative parameters, will be the transformation ratios and all this part of the process will be associated later the memory module.

Later you will psicoacústics properties are associated with visual perception parameters whose relationship will be controlled by the drive. The drive controls the flow of information input to the module relationship (Audiovisual Relation 3L) depending on the competition generated between: first, hearing and visual fixed laws; second (the modules in green), audiovisual ratio created by the relationship systematic; third relationships evaluated by the user.

The relationship systematic (Audiovisual Relation 3L) receives relations evaluated by the user and the decision taken by the drive, then creates a new relationship based on the input information and make the decision that is sent to the world of Behaviour. It also makes the decision of relations for three levels of representation: The first level is the first and most basic, and serves to relate the auditory and visual perceptions; eThe second level relates representations of emotion; The third level could relate semiotic representations as a Symbol, Icon, Index, in the future.

The Behavior module organizes and distributes the relations to be evaluated by the user by way of interaction. In next step will receive the result of the user and decide whether to send the result to Drive again to create a new proposed relationship or must be sent to the module learner to consolidate a "relative" relationship to the user.

Finally, if the user is agree with the relationship, then continues to the learning module and creates categories for relationships and sends new knowledge, establishing new rules of relationship and will competing again in the module Audio-Visual Relation 3L.

2.5 PERCEPTION

The Generation Concepts begins with the definition for levels of categorization of Sound

2.5.1 THE FIRST LEVEL

The First level, represents the most basic type of perception for connecting two states (Unconscious Perception); Visual Perception Mechanisms and the Psychoacoustic Mechanisms. The relationship between the two states represent the raw material for interconnect sounds with images, for example, the amplitude of a signal may represent the proximity of a particular object to the user, if the sound is louder, the object is nearer and larger; if the sound is softer, the object is more secluded and smaller.

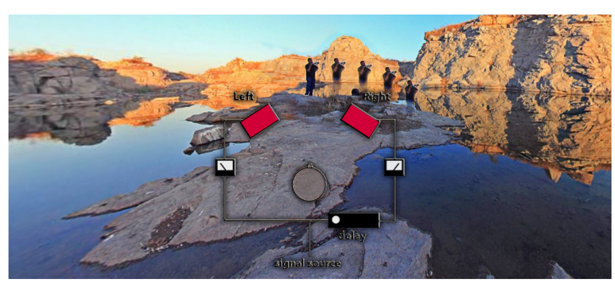

Figure 7

Another example could represent the first level is the Hass effect: If two sound sources distributed as in the Figure 7, at the same time, represents a sound source in front and in the centre of horizontal space but if you have a delay of 250ms sound source rotates 30 degrees.

It is easy to assume that this principle accompanies the movement of an animation on horizontal axis

2.5.2 THE SECOND LEVEL

Using the relationships established in the First Level, could generate a new categorization for relationships. At this level you can define motivation or emotion (Conscious Perception) for example; in music a major tonality or a minor tonality represent a state of joy or sadness, Also can visually represent states such as stress or fear, irritation, peace, etc. and will always depend on the phenomena analyzed by the system in the audio signal. This second level would work directly on the module Drive or the Behavior to create new audiovisual relationships for this second level.

2.5.3 TERTIARY LEVEL

Tertiary Level(Conscious Perception), Define the Perceptos: Transformation the World Sensories ⇒ World of Objects, then the World **Sound Sensories** ⇒ World of **Visual Objects**.

The question is: How the object is defined?. The object is the "Audio-Visual" relationship, This object must have attributes and actions and any relationship established by the system for this third level, you must follow these rules for their definition. One way to illustrate this relationship as an object, is a sound frequency, its visual representation can define a sound is above or below, even if the sound is bright or thick, light or heavy, can have performed up or down; to display the complete example use:

- Object (Sound Frequency)
  - Attributes (Qualities)
    Can a Sound be Thick or Bright?
  - Actions
    - Performed by the object
      Feel Bigger or Smaller?
    - Performed on the object (Affordances)
      It can Sound Up or Down?

The scope of the project is to get a semiotic interpretation for the third level.

2.6 ATTENTION

Attention is the process that filters the information perceived, the information to be filtered is an analysis to find psychoacoustic topics but the main work is the attention be directed to find relationships in the Audio-Visual field.

2.6.1 Attention Functions

Is easier to understand the functions of attention as; world model, the focus, and the resource manager. World model are the rules in the process of generation of concepts constituted by the three levels, starting from psychoacoustics related to visual perception; Attention is focused on the ongoing evaluation of the relationship at all levels, trying to manage the system towards better user response; The resource manager (Driver) belongs to the administration for relationships designed by the user and the system; it depends on the new learning tested by the user, where are established the new audiovisual relationships, and also depends on the second level of generation concepts through motivations and emotions.

- **Location Sensory** ⇒ Event Orientation in Space-Time (Word Modelling) ⇒ Psychoacoustic.
- **Relevance** ⇒ Signal Detection For Conscious Processing (Attention Focus) ⇒ User Evaluation.
- **Warning or Supervision State** ⇒ Prioritizing Events (Driver) ⇒ Administration of the Relationship Learning.

2.6.2 GUIDELINES TYPES

Exogenous Attention (bottom-up): The system can react exogenously with the Signal Processing and Work done by the User Input. Exogenous attention is prioritized first by the user and then by the input signal and the generated relations in the process, The Audiovisual relationships are adapted to the user, the user is constantly involved in the new generation of audiovisual relationships.

Endogenous Attention (top-down): The system can react endogenously with the internal generation of Relationships, each new relationship triggers a new chance to be evaluated by the user, the system must wait for the result by Endogenous Attention. There may be a creative process to generate new relationships from models of artificial creativity.

2.6.3 Priming

There should be a starting point to begin to establish relationships, the highest level for the generation of concepts develops attempts audiovisual association with basic phenomena. It will be the Priming phenomenon as a unconscious influence to understand Relationships, but the user is expected to make a contribution to the relationship audio-visual.

2.6.4 Salience

The system is responsive to user's relationships and final goal, pretend find a world model for audiovisual relations considering input from multiple users.

2.7 CATEGORIZATION

The system requires different types of classification; el primer tipo, proviene de analizar la señal de audio de entrada, en este proceso se identifican los fenomenos psicoacústicos. The second classifier the class relationships that are suggested to the user.

The following type of classification is organized from the experience of learning machine, entry is the user evaluation and possible relations obtained will be again classified to patronize the audiovisual association for the user. In the modulus of relationships, classified by categories of relationship types. The last classification relates in a database all possible relations established by previous users, the latter detects general patterns of relationship.

- Overall Rating:
  - Ranking events for the Input Signal.
  - Classification of Psychoacoustic Parameters.
  - Classification for the Relationships Image-Audio.
  - Classified User Inputs in relation to the Learner.
- Rating Pattern Recognition
  Grouping Classification (Categorization).
- Directly on sensory signals Signal in time
  Arrays Signals - Feature Extraction.
- Grouping (Clustering)
  Classification Problem (Classes Must Be Found).

2.8 CONCLUSIONS

This work is only a proposal to implement a cognitive system and study the first modules that were most relevant to study the relationship audiovisual.

Is necessary to thoroughly study all modules like behavior, memory, emotion, lerning, semiotics etc. to get a better definition for architecture.

It is important to the implementation to know the needs not provided theoretically; it is also appropriate work out the details to continue developing the cognitive architecture.

Defining modules perception, attention and categorization evidently emerge the modules for; behavior, memory, emotion, learning, and semiotics; which can be studied after the implementation os develop modules on perception and attention.

2.9 RECOMMENDED ARTICLES

[1]K. Yamawaki and H. Shiizuka, “Correlation of synesthesia and common recognition concerning music and color,” in Systems, Man and Cybernetics, 2004 IEEE International Conference on, 2004, vol. 1, pp. 350–357. [2]T. Miyasato, “Generation of emotion spaces based on the synesthesia phenomenon,” in Multimedia Signal Processing, 1999 IEEE 3rd Workshop on, 1999, pp. 463–467. [3]Y. Yamaguchi, K. Noda, S. Nishide, H. G. Okuno, and T. Ogata, “Learning and association of synaesthesia phenomenon using deep neural networks,” in System Integration (SII), 2013 IEEE/SICE International Symposium on, 2013, pp. 659–664. [4]G. Sing, “Mixing Audio and Animation: Contemporary Synesthesia,” IEEE Computer Graphics and Applications, vol. 21, no. 5, pp. 4–5, 2001. [5]Y. Xiaochen, S. Xiaoyong, Z. Ying, and J. Deming, “Multimodal interface design based on synaesthesia effect,” in Computer-Aided Industrial Design and Conceptual Design, 2009. CAID and CD 2009. IEEE 10th International Conference on, 2009, pp. 1379–1384. [6]A. Lee, K. W. Lai, F. T. Hung, W. L. Leung, K. K. Lam, and C. C. Leung, “Synaesthesia: multimodal modular edutainment platform development,” in Cyberworlds, 2004 International Conference on, 2004, pp. 335–342. [7]J. Villegas and M. Cohen, “Synaesthetic Music or the Ultimate Ocular Harpsichord,” 2007, pp. 523–527.

REFERENCES

[1] W. Köhler, “Gestalt psychology,” *Liveright*, vol. 2, p. 224, 1947.
[2] P. T. Maurer D and M. CJ, “The shape of boubas: Sound-shape correspondences in toddlers and adults,” *Developmental Science 9*, vol. 9 (3), p. 316–322, 2006.
[3] V. S. Ramachandran, “Synaesthesia: A window into perception, thought and language,” *Journal of Consciousness Studies*, vol. 8 (12), pp. 3–34, 2001.
[4] Y. Suzuki, D. Brungart, Y. I. ans Kazuhiro lida, D. Cabrera, and H. Kato, “Psychoacoustics - facts and models,” vol. 1, pp. 139–262, 2009.
[5] ——, “Psychoacoustics - facts and models,” vol. 1, pp. 3–138, 2009.
[6] C. Ware, “Information visualization - perception for design,” *Static and Moving Patterns - GESTALT LAWS*, vol. 3, pp. 181–196.
[7] ——, “Information visualization - perception for design,” vol. 3, pp. 450–451.
[8] H. F. E. Zwicker, “Psychoacoustics - facts and models,” vol. 3, 2006.

APPENDIX A PSYCHOACOUSTIC TOPICS

[8]

- A-weighting, a commonly used perceptual loudness transfer function
- ABX test
- Auditory illusions [4]
- Auditory scene analysis incl. 3D-sound perception, localisation [4]
- Binaural beats [4]
- Deutsch’s Scale illusion
- Equivalent rectangular bandwidth (ERB)
- Franssen effect
- Glissando illusion
- Haas effect
- Hypersonic effect
- Language processing [5]
- Levitin effect
- Musical tuning
- Noise health effects
- Octave illusion
- Pitch (music)
- Precedence effect
- Psycholinguistics [5]
- Rate-distortion theory
- Sound localization [5]
- Sound of fingernails scraping chalkboard
- Source separation
- Sound masking
- Speech recognition
- Timbre
- Tritone paradox

APPENDIX B VISUALIZATION PERCEPTION - GESTALT LAWS

- *"GESTALT LAWS [6]*
  - *Proximity*
  - *Similarity*
  - *Connectedness*
  - *Continuity*
  - *Symmetry*
  - *Closure and Common Region*
  - *Figure and Ground*
  - *More on Contours*
  - *Representing Vector Fields: Perceiving Orientation and Direction*
  - *Comparing 2D Flow Visualization Techniques*
  - *Showing Direction*
- *CHAPTER 6 [7]*

G6.01 *Place symbols and glyphs representing related information close together.*

G6.02 *When designing a grid layout of a data set, consider coding rows and/or columns using low-level visual channel properties, such as color and texture.*

G6.03 *To show relationships between entities, consider linking graphical representations of data objects using lines or ribbons of color.*

G6.04 *Consider using symmetry to make pattern comparisons easier, but be sure that the patterns to be compared are small in terms of visual angle (<1 degree horizontally and <2 degrees vertically). Symmetrical relations should be arranged on horizontal or vertical axes unless some framing pattern is used.*

G6.05 *Consider putting related information inside a closed contour. A line is adequate for regions having a simple shape. Color or texture can be used to define regions that have more complex shapes.*

G6.06 *To define multiple overlapping regions, consider using a combination of line contour, color, texture, and Cornsweet contours.*

G6.07 *Use a combination of closure, common region, and layout to ensure that data entities are represented by graphical patterns that will be perceived as figures, not ground.*

G6.08 *For vector field visualizations, use contours tangential to streamlines to reveal the orientation component.*

G6.09 *To represent flow direction in a vector field visualization, use streamlets with heads that are more distinct than tails, based on luminance contrast. A streamlet is a glyph that is elongated along a streamline and which induces a strong response in neurons sensitive to orientations tangential to the flow.*

G6.10 *For vector field visualizations, use more distinct graphical elements to show greater field strength or speed. They can be wider, longer, more contrasting, or faster moving.*

G6.11 *Consider using texture to represent continuous map variables. This is likely to be most effective where the data varies smoothly and where surface shape features are substantially larger than texture element spacing.*

G6.12 *In order to make a set of nominal coding textures distinctive, make them differ as much as possible in terms of dominant spatial frequency and orientation components. As a secondary factor, make texture elements vary in the randomness of their spacing.*

G6.13 *Use simple texture parameters, such as element size or element density, only when fewer than five ordinal steps must be reliably distinguished.*

G6.14 *To display a bivariate scalar field, consider mapping one variable to color and a second variable to variations in texture.*

G6.15 *To design textures so that quantitative values can be reliably judged, use a sequence of textures that are both visually ordered (for example, by element size or density) and designed so that each member of the sequence is distinct from the previous one in some low-level property.*

G6.16 *When using overlapping textures to separate overlapping regions in a display, avoid patterns that can lead to aliasing problems when they are combined.*

G6.17 *When using textures in combination with background colors for overlapping regions, choose lacy textures so that other data can be perceived through the gaps.*

G6.18 *When using lacy textures in combination with colors for overlapping regions, ensure luminance contrast between texture elements in the foreground and color- coded data presented in the background.*

G6.19 *To display discrete data with more than four dimensions, consider using color-enhanced generalized draftsman's plots in combination with brushing.*

G6.20 *Make every effort to standardize the mapping of data to visual patterns within and across applications.*

G6.21 *In search tasks for infrequent targets, insert retraining sessions during which targets are frequent and feedback is given regarding success or failure.*

G6.22 *When developing glyphs, use small, closed shapes to represent data entities, and use the color, shape, and size of those shapes to represent attributes of those entities.*

G6.23 *Use connecting lines, enclosure, grouping, and attachment to represent relationships between entities. The shape, color, and thickness of lines and enclosures can represent the types of relationships.*

G6.24 *As an alternative to arrows to represent directed relationships in diagrams, consider using tapered lines with the broadest end at the source node.*

G6.25 *Use closed contours, areas of texture, or areas of color to denote geographic regions. Use color, texture, or boundary style to denote the type of region.*

G6.26 *Use lines to represent paths and linear geographic features. Use line color and style to represent the type of linear feature.*

G6.27 *Use small, closed shapes to represent point entities, such as cities, that appear small on a map. Use color, shape, and size to represent attributes of these entities.*

G6.28 *Consider using a treemap to display tree structured data where it is only necessary to display the leaf nodes and where it is important to display a quantity associated with each leaf node.*

G6.29 *Consider using a node–link representation of a tree where the hierarchical structure is important, where internal (non-leaf) nodes are important, and where quantitative attributes of nodes are less important.*

G6.30 *When animation is used in a visualization, aim for motion in the range of 0.5 to 4 degrees/second of visual angle."*