inter·noise 2007 istanbul

Essentials for description of cross-sensory interaction during perception of a complex environment

Michael Haverkamp*
Ford Werke GmbH
MC1/D13
50725 Köln
Germany

ABSTRACT

It has been shown before that an assessment of sound-scapes requires integration of cross-sensory data into an evaluation process. From those first assumptions, however, it remains questionable how this information can be represented and included into an overall procedure to estimate the perceptual impact on an individual. Environments for living and working in general provide complex stimuli of various sources, like noise, light, smell, tactile information and others. The first step of perceptive signal processing is to handle stimuli of various senses separately. The perceptual system then correlates the processed attributes and provides multi-sensory integration. The aim of sensory data processing is to establish multi-sensory *perceptual objects* estimating the structures of the world exterior to the perceiving subject. Intention of the presented approach is to provide elements needed to establish a cross-sensory model which serves to estimate perceptual relevance of a given environment. The discussed essential features for a model refer to a network structure that includes perceptual phenomena (images, sounds etc.) as well as associative connections and semantic representations.

1 INTRODUCTION

It is well known from daily experience that perception of sound is influenced by stimuli of other senses. Multi-sensory properties of an environment have influence on subjective assessment of a given noise situation [1]. The appearance of visible structures and movement can significantly modify perception of sound. As an example, an unknown noise can be very annoying, until the source has been identified, e.g. a vehicle performing specific driving operations. Therefore it has to be analyzed how the multi-sensory properties of perceived objects relate to single stimuli [2]. Especially the integration of all stimuli apparent at one time defines an overall ambience (*atmosphere*) into which a single perception will be fitted. Figure 1 gives an example of an ambience which contains audible and visible objects, with some objects also showing movement.

Currently only few approaches are available to include single aspects of the multi-sensory content of sound stimuli into assessment, like e.g. associative aspects [3]. A complete and satisfying analysis, however, requires much more effort. While there are more than only single perceptual strategies active to provide a multi-dimensional and meaningful representation of the physical world surrounding the individual, it has to be considered that an comprehensive analysis of an environment can not only be based on simple relations.

The approach presented here tries to collect parameters and processes of multi-sensory coupling which are expected to provide a fundament of more sophisticated assessment procedures and computerized analysis of complex environments.

* Email address: mhaverka@ford.com

2 PERCEPTUAL OBJECTS

In general, the perceptual system tends to provide multi-sensory models of physical objects. These models are needed by the individual to interpret his environment and to coordinate his actions, while it is surrounded with objects which physical nature cannot be accessed directly. As a result of perception and cross-sensory integration, an image occurs in consciousness. This image can show aspects of vision, audition or of any other sensory channel. In contrary to the *physical object*, these subjective representations are here named *perceptual objects*. A sensory channel will further on be described as *modality*.

A subjective representation of a physical object is based on a sensory hypothesis, generating a perceptual object that in some ways correlates to its physical source. While physical objects always provide various stimuli, perceptual objects usually appear to be multi-sensory, i.e. they contain auditory, visual, tactile and other data. Only if stimuli of a single sensory channel are presented which are unknown to the individual, a first approach is made by generating perceptual objects of this single modality, e.g. auditory or visual. **It therefore can be stated that in case of known stimuli assessment of perceived noise is always related to multi-sensory perceptual objects.** Unknown signals, however, are meaningless and not attributed to a known source, which otherwise would provide multi-sensory stimuli.

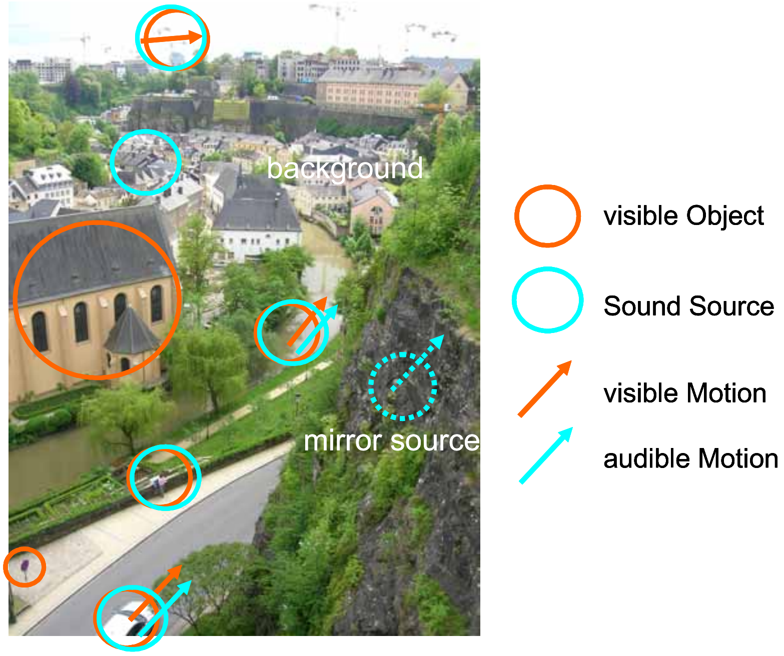

Fig. 1: A complex environment contains various sources for perceptual stimuli:
Environment at Luxembourg City, 2006

A multi-sensory perceptual object can be understood as a cluster of perceptual objects of single modalities. The data of different modalities can be coupled by various strategies [2]. The main *intuitive strategies* (*coupling features*) are:

analogies = correlations of single features/attributes,
iconic coupling (concrete association) = identification of sources (of stimuli)
symbolic connections = semantic correlations by analysis of meaning

Each main intuitive strategy splits into a variety of mechanisms, e.g. analogies can be made up in-between basic attributes, like brightness or roughness, or can consider movement, shape, emotion and many other aspects.

The listed main strategies of interaction are interpreted intuitively and do not need conscious analysis. Other correlations can be constructed with reference to known physical properties, e.g. the frequency of color light can be related to pitch frequency of a given pure tone by means of appropriate calculations. Common ways to provide constructions can easily be found by programming specific algorithms, as used by audio-visual media players and automated light-shows. If the perceiving subject, however, cannot find any correlations by use of the abovementioned intuitive strategies, it will not be able to find any match of data provided by different senses.

It is assumed that the listed main strategies of interaction are acting in-parallel, showing mechanisms that initially are independently from each other. Therefore, for the first step of perception the analysis can separately refer to those strategies. The next step then must compare the correlations and integrate the results to provide a map of correlations which is consistent, not including any contradictions. Assumptions regarding this process of integration are discussed below.

Shaping of perceptual objects from a variety of data requires integration of all perceived elements. This includes various mechanisms of grouping and segregation. First approaches which are also valid today have been provided by Gestalt-psychology [4]. Recently, theories of scene analysis have added essential insight into integration processes [5].

The quality of a perceptual object is only experienced by the perceiving subject itself. The various qualitative aspects inherent to a specific perceptional object are named *qualia*. Perceptual objects are not only sensed while stimuli are present. They are also recalled from memory. The stored mnemonic information is used to completely remind a known object or to complete it in case stimuli are only partly available.

The aim of the perceptual system is to identify physical objects by generating those multi-sensory perceptual objects. If this object is known and represented in memory, its cross-sensory features can be recalled by stimulation of only one single modality. Learning to handle physical objects of daily life requires testing of all sensory properties, like vision, audition, smell, taste, surface structure, hardness and many others. The learning process can easily be observed at a baby handling a toy. It has a look, immediately grasps it, hits it towards another object to experience its sound, puts it into its mouth. After a while, it has learned that a specific sound indicates a hard or a soft surface, a heavy or lightweight object etc. It then is enabled to remind the perceptual object of a toy by hearing the sound, and this recalls all cross-sensory properties which have been experienced before.

From these observations it must be concluded that the task of a model that estimates the subjective relevance of a given environment shall not only refer to present stimuli, but must refer to the multi-sensory perceptual objects stored in memory, which are activated by the stimuli.

Parallel processing is a basic feature of the brain [6]. It can be found on all levels of neuronal activity, from the interaction of single neurons with its inhibitions and amplifications up to the binding activity of complete cortical domains.

Remark: Consideration of the described three main strategies of cross-sensory coupling implies that contextual information is processed within each strategy.

The analysis of all single elements and attributes of a complex sensory environment can be understood by means of a multi-modal network, as sketched in Fig. 2. This network can contain both iconic (concrete associative) and symbolic elements. The iconic elements are stored in memory as shapes/images of sensory perception, i.e. as visual image, sound, odor etc. The number and grade of cross-linking of elements is determined by the level of experience (number and intensity of perceptional events) and by the mental und emotional involvement of the perceiving subject.

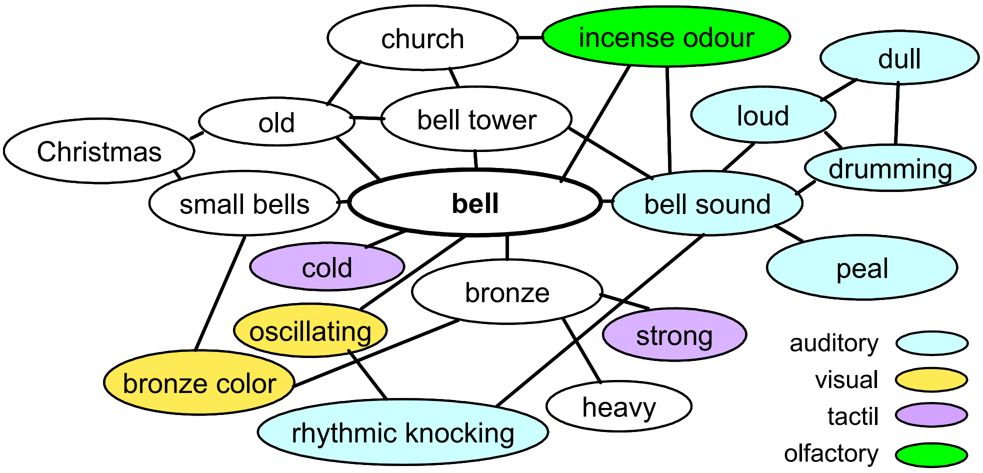

Fig. 2: Multi-modal network – Example: "bell"

With reference to the aforementioned perceptual objects, they can be understood as connected clusters of attributes, for which a certain probability exists that they belong to a specific physical object (or term/item). If this cluster is intuitively integrated within a single modality, it defines a perceptual object of one sensory channel. Attributes of single modalities can be connected via analogies, i.e. correlation functions. If grouped to perceptual objects of single modalities, those objects can be connected via iconic or symbolic coupling.

3 FOREGROUND-BACKGROUND ALLOCATION

The intuitive analysis of an environment must also include selection of prominent perceptual objects, which claim attention and communicate meaning [7]. On the other hand, less important objects are merged to a broad background that mainly contains atmospheric information.

Typical foreground signals are sharp in their temporal, spatial and/or spectral properties. It is significant that most of Schafers examples of sound-marks are noise sources for localization, like fog horns, or for temporal information, like church bells or a cannon indicating specific hours [8]. Typical natural sound-marks for localization are used by animals, like a wolfs howling, chirping of a grasshopper or singing of birds. A visible object that serves as a land-mark must also show high visibility and enable exact bearing, like a church tower or a sharp mountain. Examples for sharp spectral properties are a traffic light with its pure colours and typical emergency sounds, like sirens and horns.

On the other hand, typical signals assigned to the background do not show sharp localisation, are continuous in time and have broad spectral properties. Background noise of a city or a forest is equally distributed to all directions. Human hearing accustoms to continuous signals, which are the less capable to contain specific information. Statistics shows that a sound source which contains many processes with small influence on the whole equals a broad-band random noise. Because of its different spectral properties and less temporal fluctuation, a distant highway often causes less annoyance than a small country road close to the listener. Both foreground and background stimuli contain specific information as well as atmospheric content.

Fig. 3 demonstrates that simple modifications of background properties or land-marks can fundamentally change the visual appearance and ambience of an environment. This observation is valid for landscape analogies of all modalities, e.g. sound and tactile environments show a similar behavior. Very few elements can enable appropriate identification of a sound-scape. This fact is used for movie sound design [9].

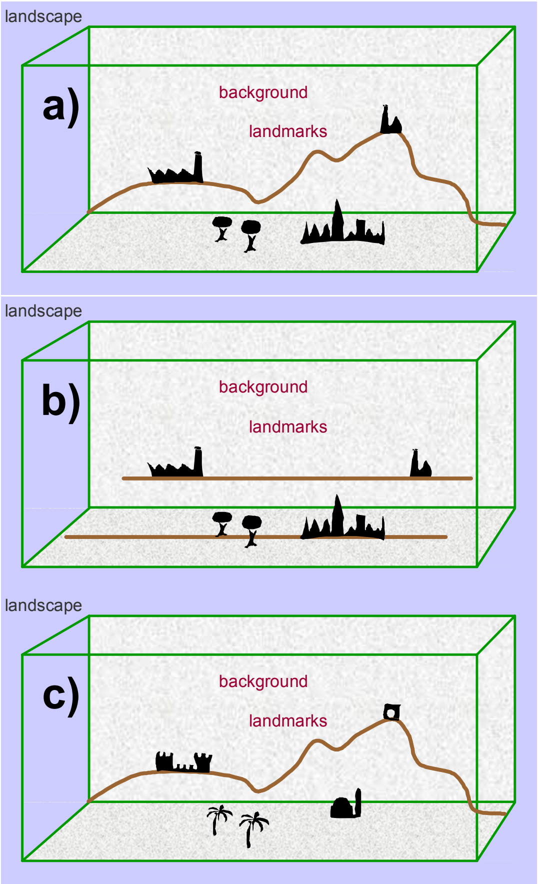

Fig. 3: Similar to a sound-scape, the properties of a land-scape are defined by few elements. The scene is divided into background (symbolised by brown lines) and land-marks of the foreground (black).
b) modification of the background immediately changes the ambience
c) modification of the foreground changes the ambience and can also change the semantic content connected to the land-marks (e.g. the cultural affiliation).

It must be clear that changes of specific parameters of auditory or visual events can only be compared if the identity of perceptional objects is not change.

In example, the pass-by-noise of a single vehicle is perceived as a quite different object than the noise of hundreds of vehicles on a motorway. In the first case, the sound properties depend on the specific operation of the vehicle. Its sound shows a specific onset phase and an certain decay phase with different sound quality due to the specific directivity of noise propagation and the Doppler phenomenon. The noise includes dynamic features which indicate the specific driving operation, like accelerating engine speed, tire squeal and others.

Compared to this, the merging of many vehicle contributions to make up the sound of a motorway generates a single perceptual object with absolute different object qualities: The dynamic features merge into a constant noise. Hence the perceived signals change from typical foreground properties (limited in temporal, spectral and spatial domain) to background properties, given by temporal constancy, broad-band spectrum and wide area of sound sources. If one identical criterion (e.g. L_{eq}) is applied to both a single vehicles noise and a widely used motor way, comparability is limited while assessment refers to absolutely different perceptual objects.

The identity of perceptual objects is an essential precondition of reliable assessment.

4 COUPLING STRATEGIES

While various strategies of cross-modal coupling are relevant for shaping of multi-sensory perceptual objects, a multi-modal network as shown in fig. 2 shall be split up to clarify the different processes.

4.1 Cross-sensory Analogies

Cross-sensory analogies refer to the capability of the perceptional system to detect correlations of specific attributes and to analyze them for identification of physical objects and atmospheric features [2, 10].The analysis of analogies can refer to:

- generic attibutes (intensity, sharpness, brightness ...)
- motion (*straight*, *rotational*, *irregular*, *expanding* ...)
- body perception (*tense*, *relaxed*, *floating* ...)
- emotion (*calm*, *troubled*, *angry* ...)

Correlations of generic attributes have been comprehensively discussed by Stevens [11]. For the aspects of motion as a visual and auditory attribute see [12] and [13].

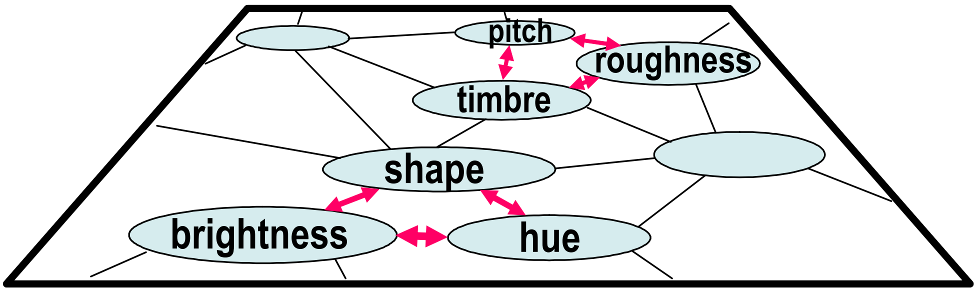

Fig. 4: Correlation of cross-sensory analogies, exemplified in case of perceived visual and auditory attributes

The capability to detect cross-sensory correlations as well as correlations of various attributes within one sensory space is an essential feature of the perceptional system. Without the ability of evaluating analogies, perception would not be possible because the build-up of perceptual objects would not be enabled. A jingling bell can only be recognized if the jingling noise (auditory space) is connected to the image of the bell (visual space). If the necessary integration process fails, only a noise with unknown source is detected in parallel to a moving visual object that cannot be determined as sounding or not sounding. Identification of physical objects and analysis of the complex environment which surrounds the individual is not imaginable without determining correlations.

Analogies are capable to consolidate unknown or unexpected perceptual objects by correlating the perceived attributes. A sound source localized in a specific angle and distance will be coupled to a physical object seen near its location. A connection will also be presumed in case of different location, but with accurate temporal correlation (*synchronicity*). Therefore spatial and temporal correlations must be primarily be considered. In example, a heard speech can be allocated to a specific person via synchronic sound and motion of the mouth, additionally supported by correlated visual and auditory localization.

4.2 Iconic Coupling / Concrete Association

This strategy to establish cross-modal connections is based on associations suitable to identify a known physical object. A single stimulus can refer to a multi-modal perceptual object stored in memory. Thus a specific sound stimulus can evoke imagination of the sound source with all of its cross-sensory attributes, if the variety of properties was experienced before. In example, the sound of the siren of an ambulance refers to the image of emergency light and of the whole vehicle.

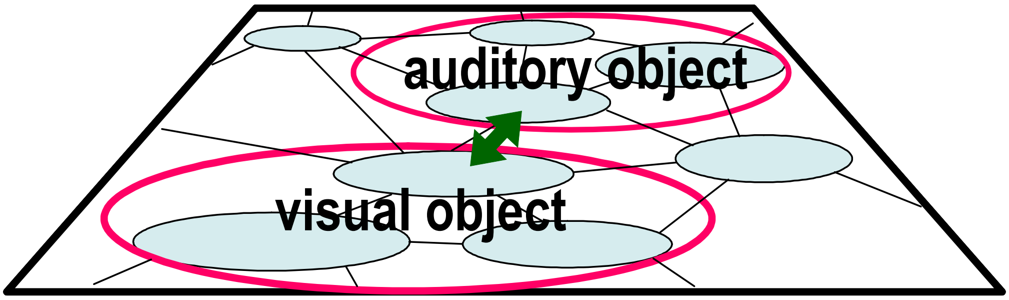

Fig. 5: perceptual objects of single modalities, grouped from the visual and auditory attributes of fig. 4 as base of iconic coupling

Listening to music often evokes imagination of landscapes or interiors which fit to the associative content of the music or a comparable atmosphere. Onomato-poeia in speech and music is a common application of iconic features, while the imitation of natural sound generates an intuitive connection to multi-sensory objects and to the atmosphere of an environment.

While iconic coupling refers to objects in memory, it is based on learning and experience of the subject. It therefore depends on living environment and cultural background of an individual. If an automated analysis of an environment is intended, this information must be provided by a database which is consistently prepared regarding a given context.

4.3 Semantic relations

Stimuli of a specific modality can also refer to symbolic (semantic) codes, which e.g. are given in the visual modality by signals and logos, often based on aspects of ancient heraldry. Those symbols can only be understood if sender and recipient of a message are based on the same context. Therefore the functionality of signs and symbols is limited to specific aspects of culture and era.

Synaesthetic terms as a special kind of metaphor are an important element of literature,

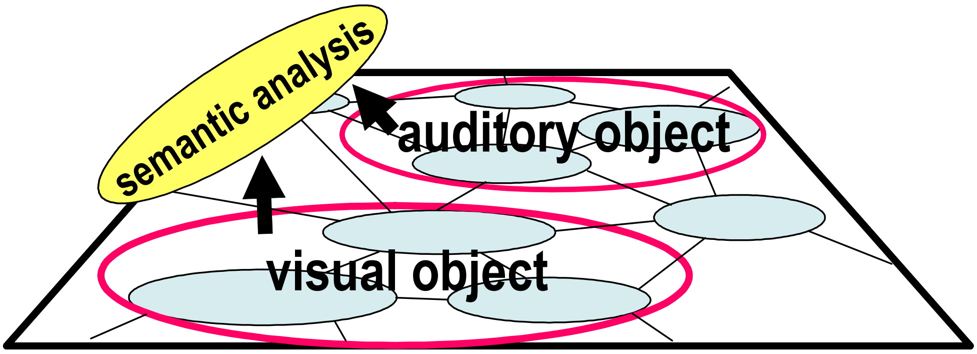

Fig. 6: additional relations of sensory objects are established using a semantic analysis, enabling multi-sensory grouping by correlations of meaning as base of symbolic coupling

especially of lyrics. Synaesthetic metaphors are also applied for description of sensations during psychophysical experiments. In example, a sound can be described as *warm*, *bright* or *heavy*. Those terms are primarily conscious constructions, but need the recipient's capability of imagination via cross-sensory analogy or association. Semantic relations refer to a known source of stimuli and to a known semantic content. Speech that is understood includes semantic information as well as iconic features suitable to identify the source, e.g. a person. Speech of a language unknown to a subject, however, only refers to iconic coupling and cross-modal correlations.

5 ESTIMATION OF THE OVERALL EFFECT

The perceptual system collects the information of all modalities and performs sophisticated integration processes to merge the naturally incomplete data into a multi-sensory image of the world. The individual's world includes exterior as well as interior body perception. The aim of the integration processes is to generate subjective representations of objects which are free of contradictions and gaps. Those representations are then base of final assessments of the subject's existential orientation, which includes well-being, annoyance, danger etc. Therefore an analysis of a complex environment for estimation of its effect on the quality of life must take into account various processes and steps of integration.

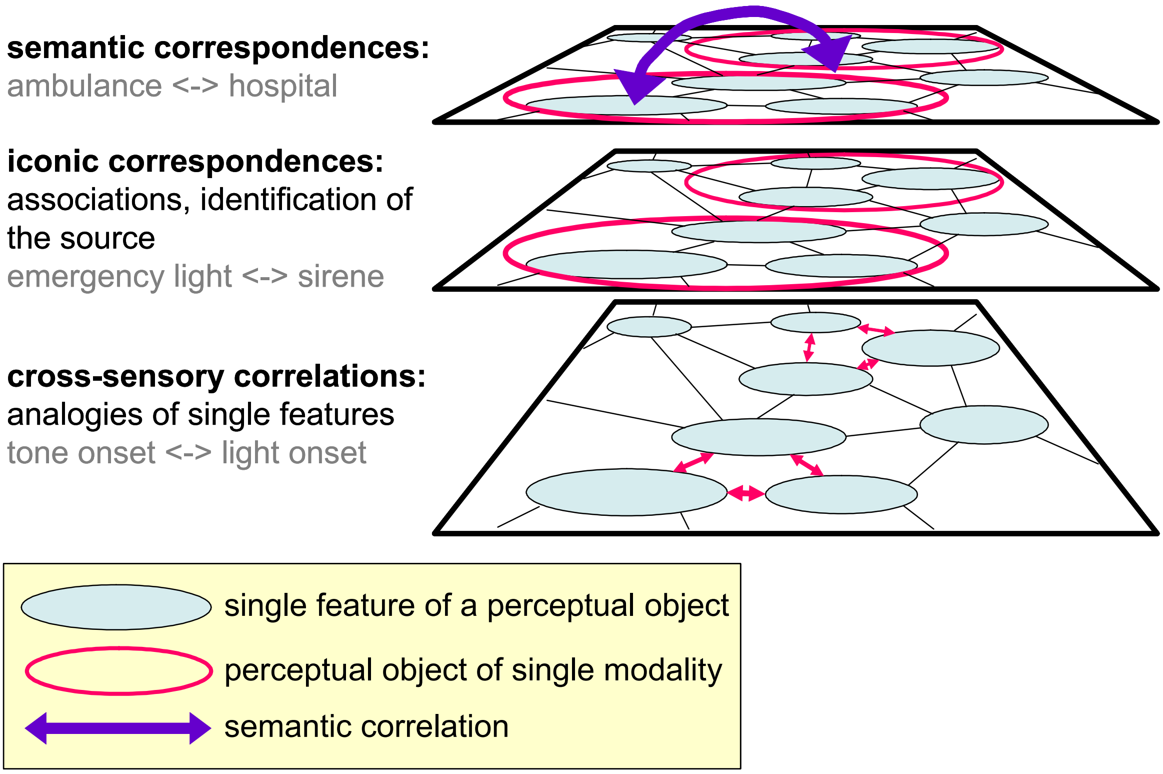
semantic correspondences:
ambulance <-> hospital
iconic correspondences:
associations, identification of the source
emergency light <-> sirene
cross-sensory correlations:
analogies of single features
tone onset <-> light onset
single feature of a perceptual object
perceptual object of single modality
semantic correlation

Fig. 7: Three-layer model for in-parallel simulation of cross-modal analogies, iconic coupling and semantic correlations.

The approach presented here is based on the fact that the perceptional system is able to analyze analogies, iconic and semantic strategies of cross-modal coupling in-parallel. Therefore it appears to be reasonable to define a network with 3 layers for multi-sensory analysis of a given situation, as shown in fig. 7. In the first step, it implies that characteristic attributes of perceptual objects are clearly perceived and segregated from each other. Then, the allocation to fore- or background shall clarify whether objects have a prominent relevance or are integrated into a variety of less important objects which as a whole contribute to the atmospheric quality of the environment. On the level of analogy, correlations of single properties are evaluated. This must refer to uncertainties and inaccuracy of stimuli as well as of fuzziness of perception. In example the *localization blur* and temporal delays due to different propagation velocities of sound and light have to be compensated. Sound reflections and light refractions are possible sources of misperception which has to be avoided.

Specific thresholds for integration of deviations must be taken into account, e.g. the *echo threshold* as the minimum delay between two pulses necessary to enable segregation by the auditory system. In general, correlations help to integrate perceived features to multi-sensory perceptual objects and to associated processes.

As the next step, the analysis of iconic content identifies and connects attributes which are recognized as elements of known perceptual objects. Even a rudimentary stimulus within one modality is capable to initiate the reconstruction of a complex multi-sensory perceptual object by adding all features stored in memory. The main problem of an assessment process is that a variety of features has to be considered that is only available in memory of the perceiving individual. The set of subjective qualities (*qualia*) related to a stimulus also depends on past experience and external context. Within a known cultural and environmental context, however, it seems to be feasible to collect a data base that serves to add the relevant information that persists in memory of the persons who's overall feeling and reaction has to be estimated. Simple psycho-acoustic parameters fail since they refer to fixed relations of perception and instantaneous stimuli.

On the semantic level, perceptual objects are coupled according to additional information which is learned separately, like e.g. a language, traffic signs, philosophical, political or religious interpretation and thinking.

A basic hierarchy of the three strategies can be stated. Unknown perceptual objects can be loaded with information about the physical source and meaning by correlation analysis of their cross-sensory features and behavior. Well known objects are already defined by iconic coupling during presentation of single features. Meaning can be added by observation of known objects and by learning of additional, abstract information.

In example, a view on a driving car provides analogies of the visualized movement and dynamic features of the sound. With this experience, movement and direction of the vehicle can be reconstructed intuitively by only listening to its sound. With knowledge of a semantic system of traffic rules, and with respect to the given context, the cross-sensory data can be a base of assessment about compliance with legal requirements, danger, emotional situation of the driver and others.

The parallel processing of various strategies can thus gain independent results which can be contradictory or similar. Friendly sentences e.g. can be vocalized with aggressive facial expression, or aggressive sentences can be supported by slight body movement and gentle voice. Different strategies are used by the perceptional system to eliminate those contradictions. If visual and auditory signals contain different data, the result can be dominated by one sensory channel, while the other is inhibited. *Ventriloquists* make use of the dominance of the visual modality: the mimic action of a doll indicates that the source for its speech is at its mouth, while information about the location of the true sound source is suppressed. Another possibility is that contradictory sensory information is merged to shape a different result. This has been exemplified by means of the *McGurk-effect*, where the mimic action of a speaker modifies the sound of syllables [14]. If contradictions of data provided by different modalities cannot be matched, this can cause negative feelings like indisposition or *cognitive dissonance*. The fact that this attracts specific attention of the subject is nowadays well known and used by print and screen advertising.

6 CONCLUSION

As result of the described considerations, an appropriate, e.g. computerized analysis of the perceptual relevance of an environment must include the following essential steps:

1) integration and segregation of attributes to shape perceptual objects in all modalities, e.g. shape auditory and visual objects
2) rate perceptual objects in terms of foreground - background determination, with respect to appropriate thresholds
3) cross-sensory correlations of basic attributes (e.g. pitch, timbre, hue), based on certain probabilities and threshold assumptions
4) analyze iconic content to shape multi-sensory perceptual objects by means of an expert system, which includes data relevant for source identification
5) analyze semantic content to establish cross-modal connection of meaning, also based on an expert system
6) interpret atmospheric content of all perceptual objects
7) final integration and concluding assessment, based on probabilities of relevance of each coupling strategy, including processes for minimization of contradictions

7 REFERENCES

[1] Koji Abe, Kenji Ozawa, Yôiti Suzuki and Toshio Sone, Comparison of the effect of verbal versus visual information about sound sources on the perception of environmental sounds, Acta Acustica united with Acustica, Vol. 92 (2006), p. 51-60

[2] Michael Haverkamp, Audio-Visual Coupling and Perception of Sound-Scapes, Proceedings of Joint congress CFA/DAGA'04, Oldenburg, DEGA, (2004) p. 365-366

[3] Kenji Furihata, Takesaburo Yanagisawa, David K. Asano and Kazumasa Yamamoto, Development of an experimental noise annoyance meter, Acta Acustica united with Acustica,. Vol. 93 (2007), p. 73-83

[4] Heinz Werner, Intermodale Qualitäten, in: Handbuch der Psychologie, Bd. 1 (Göttingen, Hogrefe, 1966, p. 278-303)

[5] Albert S. Bregman, Auditory scene analysis, The perceptual organization of sound, (Cambridge, Massachusetts: MIT Press, ^{2}1999)

[6] Christoph von Campenhausen, Die Sinne des Menschen (Stuttgart: Thieme, ^{2}1993)

[7] Murray Schafer, The tuning of the world (Toronto: McCelland and Steward, 1977)

[8] The world soundscape project, Simon Fraser University, The Vancouver Soundscape 1973 & Soundscape Vancouver 1996. 2 CD (Cambridge Street Records)

[9] Barbara Flückiger, Sound Design, Die virtuelle Klangwelt des Films (Marburg, Schüren, ^{2}2002)

[10] Michael Haverkamp, Visualization of synaesthetic experience during the early 20th century, Internationalen Conferenz on Synaesthesia, Hannover, 2003. published on *http://www.michaelhaverkamp.mynetcologne.de/synaesthesie_engl.htm*

[11] Staley S. Stevens, The psychophysics of sensory function, in: Rosenbith, W.A. (Ed.), Sensory communication (Cambridge: M.I.T. Press 1961)

[12] Patrick Shove and Bruno H. Repp, Musical motion and performance, in Rink, J. (ed.): The practice of performance (Cambridge University Press, 1995)

[13] Alexander Truslit, Gestaltung und Bewegung in der Musik (Berlin, Christian Friedrich Vieweg, 1938)

[14] H. McGurk and J. MacDonald, Hearing lips and seeing voices, Nature 264, 1976, p. 746-748