

Cognition, Ballistics and the Spatial Environment Engine

Robert Reams, AES Member

In previous papers an engine for the computation of a real time automotive “spatial environment” was proposed. It was determined that a description of image cognition factors and timing was needed. This note presents performance criteria based on the ballistics limitations of human aural cognition as a function of the environment.

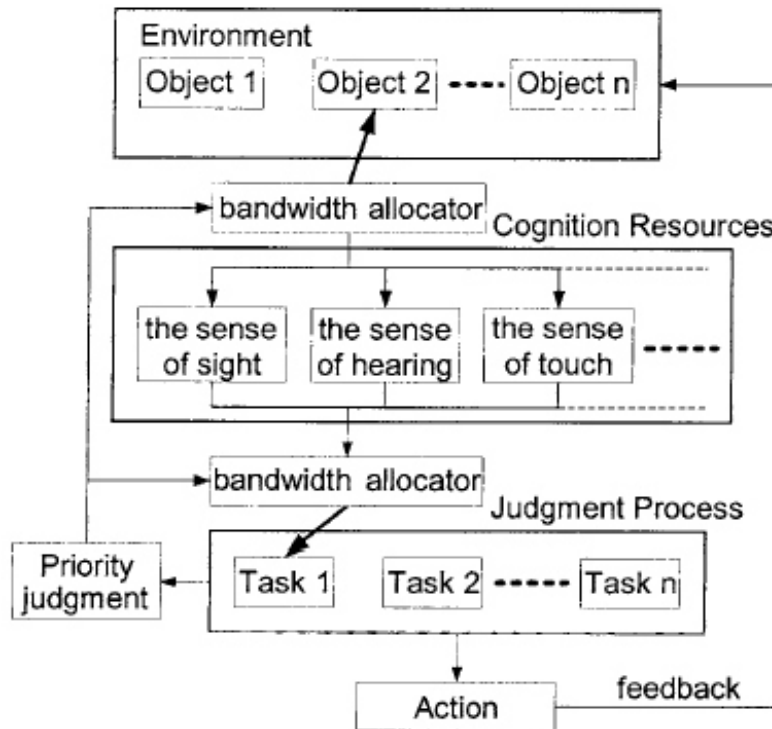
RECOGNITION RESOURCE SHARING MODEL

It is said that the majority of the recognition process utilizes vision. In instances where vision is used, such as, television/movie viewing, video gaming or driving, it can be assumed that the majority of information needed for interacting with the environment is derived from visual information. Let us consider for a moment the interaction between the visual and auditory senses. People usually recognize an object through both visual and auditory senses. Based on this recognition the person executes a judgment process. However, there are cases whereby, when a person is absorbed in something, he or she does not notice when someone calls him or her, or when a person is concentrating on music or a radio program, he or she is not aware what is happening around him or her. This can be generalized as an interruption by one recognizing resource to a person who is concentrating on a certain object using another recognizing resource. The above mentioned cases seem to indicate that it is difficult to concentrate on object A through the visual sense and, at the same time, recognize object B through the auditory sense. We therefore establish the following model for the allocation of information processing capability.

When a person has to interact with multiple stimulus, he or she allocates his or her visual or aural recognition resources according to the priorities of tasks determined in the judgment process

The schematic diagram shows a model depicting the above concept. In this model, the bandwidth allocator allocates the information source (object to be recognized) to each of the tasks, and associates the object with the appropriate recognition resource and with the judgment process of each task. The operation of the bandwidth allocator reflects the priority of each task determined in the judgment process. A task with higher priority will get more bandwidth. If no task has a particularly high priority, recognition resource bandwidths are allocated uniformly. In contrast, if a task with high priority arises, that task is allotted more bandwidth. As a result, the utilization of the recognition resources not required by that task decreases. This feedback of the priority as

determined in the judgment process results in the interaction with the surrounding environment being reflected in the bandwidth allocation in real time.



METHOD OF ESTIMATING INFORMATION PROCESSING CAPABILITY AND WORKLOAD

Recognition models in the past focused on visual recognition in cases where only dominant (P_F) and secondary (P_B) visual tasks existed. An extension of the model to include the aural recognition is needed. If aural recognition resource is included, the information processing capability allocated to visual recognition is reduced to the extent resources are allocated to aural recognition. If aural recognition takes up all the information processing capability, the line-of-vision may move about indeterminately, and it may be impossible to focus on a specific information source. Let P_H be the information processing capability allocated to aural recognition. Then, the model may be extended to:

$$[1] \quad P_T = P_F + P_B + P_H = \text{const.}$$

This expression shows that, if the information processing capability is allocated to aural recognition, the time allocated to visual recognition per unit time is reduced. Let TR_S be the proportion of visual attention time in a driving task, for which only visual recognition is required, and TR_H be the proportion of visual attention time in a situation where aural recognition is also

required. Then, the information processing capability P_H allocated to aural recognition can be expressed as:

$$[2] \quad P_H = C_R (TR_S - TR_H)$$

In other words, if the proportion of visual attention time in a visual interaction-only-task situation is taken as the reference, the information processing capability allocated to aural recognition is proportional to the reduction from the reference. C_R is a constant of proportionality.

EXTENSION OF THE METHOD OF ESTIMATING THE INFORMATION PROCESSING WORKLOAD

In a task in which only visual recognition is required, the higher the information processing workload, the higher the degree of visual attention paid, and the smaller the line-of-vision movement. This reduction in the line-of-vision movement can be considered to correspond to the workload. However, if aural recognition is also required, information processing is expended on aural recognition, leaving visual recognition in an idle state, which can increase the line-of-vision movement. Let $S(t)$ be the vector indicating the position of the line-of-vision at time t , L_H be the aural recognition-dependent workload, and L_S be the visual recognition-dependent workload:

$$[3] \quad \frac{1}{L_S} + L_H = \int S(t) dt$$

Let D_S be the summation of sight angle differentials of the line-of-vision movement when only visual recognition is required in an interactive viewing situation, and D_H be the summation of sight angle differentials of the line-of-vision movement when aural recognition is also required. Then, as in the allocation of information processing capability, the aural recognition-dependent information processing workload L_H can be expressed as:

$$[4] \quad L_H = C_D (D_H - D_S)$$

In other words, if the summation of sight angle differentials of the line-of-vision movement in a driving-task-only situation is taken as the reference, the aural recognition dependent information processing capability is proportional to the increase from the reference. C_D is a constant of proportionality.

Allocation of information processing capability and its relation to workload. According to the model proposed in previous section, if the workload on aural recognition increases, the information processing capability allocated to aural recognition increases as well. This is expressed by

$$[5] \quad P_T \propto L_H$$

Substituting Expressions [2] and [4] in [5], we get

$$[6] \quad TR_S - TR_H \propto D_H - D_S$$

In other words, if the value in a interactive viewing situation is taken as the reference, the degree of reduction in the proportion of visual attention time is proportional to the increase in the summation of sight angle differentials of the line-of-vision movement.

EFFECTS OF THE AURAL RECOGNITION DEPENDENT INFORMATION PROCESSING CAPABILITY

Studies have shown that the reduction in the length of line-of-vision movements validates that the existence of *an aural recognition-dependent workload leaves visual recognition idle*. The summation of sight angle differentials of line-of-vision movements per unit time generally agrees with Equation (4). Many subjects showed a larger variance from normal interactive visual tasks in the scenarios where aural pattern recognition and interpretation were involved. In general, it can be observed that the line-of-vision movements tend to increase as the aural recognition workload increases. Tests reveal that there is no significant difference in mono music-listening tasks and short memorization tasks, but that there is a significant difference in complex pattern interpretation tasks, long memorization tasks and aural interaction tasks. Comprehension, appreciation or emotional response may explain this difference.

ALLOCATION OF THE INFORMATION PROCESSING CAPABILITY AND ITS RELATION TO WORKLOAD

The information processing capability allocated to aural cognition is proportional to the aural cognition-dependent information processing workload. The decrease in the sight-fix time and the increase in the summation of sight angle differentials of the line-of-vision movements is an effective indicator of the two characteristics.

To develop an effective aural spatial environment for any interactive application, it is necessary to

develop a method for the quantitative estimation of the workload imposed on the consumer by multiple stimulus. Focusing on information processing capability, this is a model for the allocation of recognition resources in cases where an aural recognition task is added to the interactive visual task, which is normally dictated by visual recognition. The bandwidth allocator controls, according to the priority determined by the judgment process, the connection to the recognition resources appropriate for the object to be recognized, and the connection of recognition resources to the judgment processes. This is a method of estimating the information processing capability using visual recognition information and an expression that describes the relationship between the information processing capability and the visual recognition information.

Studies have further shown (through the measurements of line-of-vision movements) that the imposition of secondary aural-recognition tasks on primary cognition

- a) reduce the information processing capability allocated to cognition
- b) slow the tasks invoking pattern recognition and comprehension/appreciation

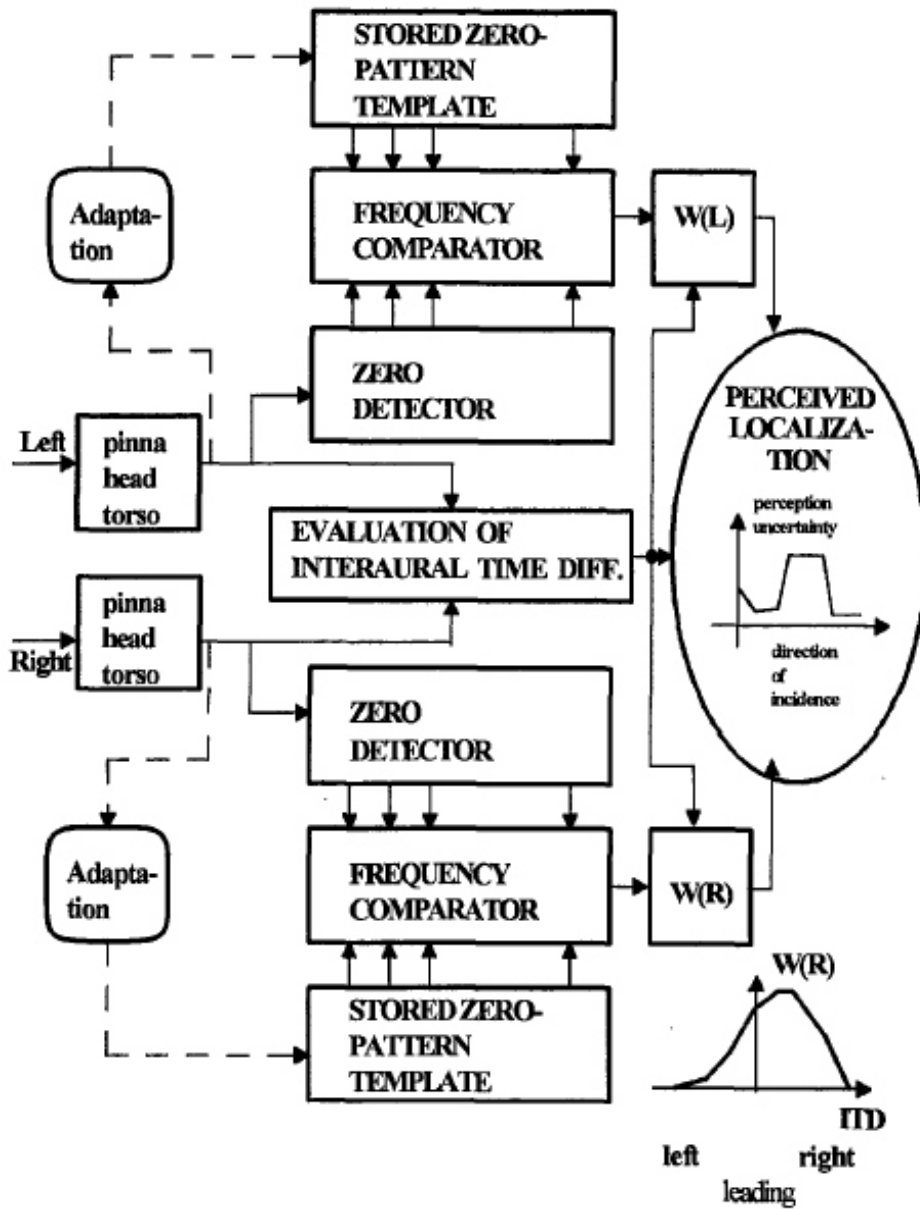
This demonstrates a clear-cut relationship to the number of simultaneous perceptual objects and the ballistics of cognition. The increase of perceptual objects modifies Equation [1] and proportionally increases perception complexity with a commensurate relative slowing in the aural cognition ability.

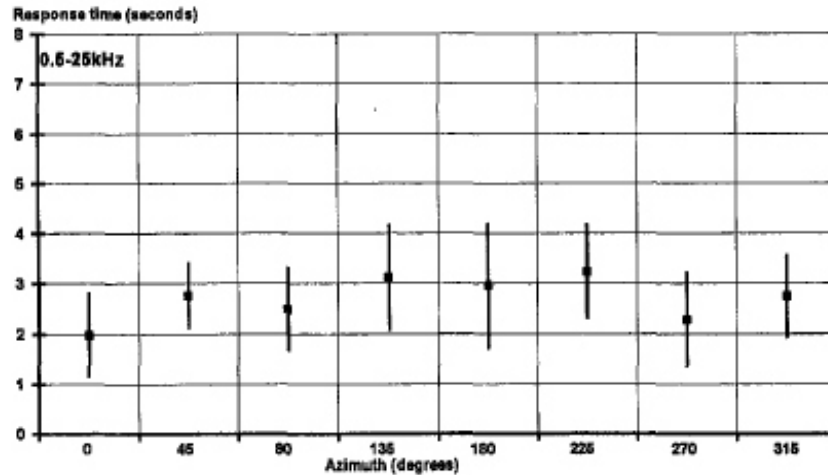
IMAGE COGNITION BALLISTICS

The nature of the localization cues has been the subject of intensive research over the years. An up to-date summary of sound localization is given by Wightman and Kistler and by Blauert. A schematic model of the human auditory localization mechanism based on the median plane localization model proposed by Blauert describes sound localization is shown.

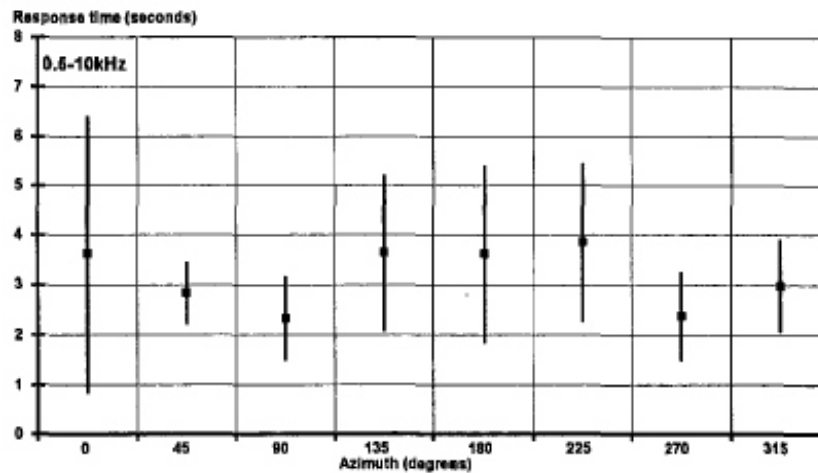
It consists basically of two parts: 1) a binaural processing mechanism which evaluates the interaural time difference between the two ear-signals and hence determines the specific cone of confusion containing the sound source and 2) two monaural pattern recognition mechanisms which determine the position of the sound source within the specific cone of confusion determined by part 1 of the model. The detailed nature of the pattern recognition mechanism is still uncertain although some of the characteristic cues utilized by the mechanism has been known for a long time. These cues consist of patterns of notches in the head related transfer functions (HRTFs). At least two of these (termed the lower frontal notch (LFN) and upper frontal notch (UFN) in the figure) have been shown previously to contain important localization information. It is shown in the present investigation that information above 10kHz is also utilized by listeners with normal hearing in the localization process.

Auditory localization is characterized by a dominance and focus mechanism that may be behaviorally described as a common Heisenberg stand-off: the more spectrum the auditory system has to work with, the faster it can "harden" an image into a stationary local. This mechanism is further impacted by the work load of the aural recognition system as described above.

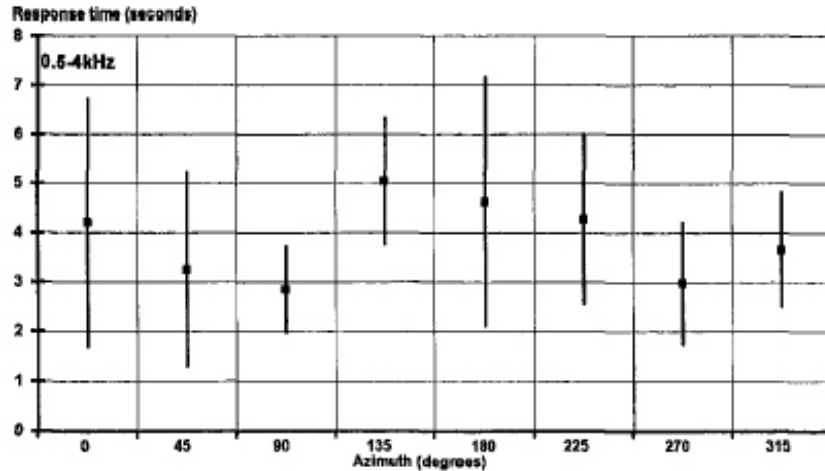




Response time (mean +/- st.dev.) obtained with signals of full bandwidth (0.5-25kHz) as a function of direction of incidence. The response time is practically independent of direction of incidence although with a slight tendency to an increased response time at backward sound incidence.



Response time (mean +/- st.dev.) obtained with 0.5-10kHz signals as a function of direction of incidence. Comparing this figure with fig. 12 (0.5-25kHz) it can be seen, that frontal localisation (0 deg. azimuth) has become a more difficult task as indicated by the significantly increased mean response time. The increased standard deviation for this sound incidence indicates that some subjects find the task considerably more difficult while others do not.



Response time (mean +/- st.dev.) obtained with 0.5-4kHz signals as a function of direction of incidence. Comparing this figure with fig. 12 and 13 a further increase of mean response time is observed. The increased standard deviation for practically all direction of sound incidence indicates that some subjects finds the task considerably more difficult while others do not.

CONCLUSION

As can be seen from the above three charts, with only one aural stimulus (no visual stimulus) image localization is accomplished in periods from around 2 to 4 seconds. It must be emphasized that these times will be further extended by additional aural image components and visual stimulus as described by the work load constant. Aural stimulus content, when restricted to 2D lateral and depth cues demonstrate much faster image localization. Generalized direction (with non-repetitious aural objects) is on the order of 100ms with fairly precise hardening within 1 second. 2D aural renderings consisting of one dominant and two recessive objects (with visual stimulus consisting of one active object) yielded generalized localization within 300ms with hardening within 1 second for the recessive objects.