7 - 12 SEPTEMBER 2014, KRAKÓW
FORUM ACUSTICUM

Evaluating the salience of auditory events through eyes

Shigeto Furukawa, Hsin-I Liao, Shunsuke Kidani, Makoto Yoneya, Makio Kashino
NTT Communication Science Laboratories, NTT Corporation, Kanagawa, Japan.

Summary
In complex acoustical environments, some sounds are more salient (or attention-capturing) than others. Experiments were conducted to explore the extent to which the auditory salience can be observable as a pupil dilation response. Stimuli were a sequence of repeating “standard” sounds, which were intermittently replaced with low probability “oddball” sounds. The presentation of infrequent oddball sound was considered to be a salient event. Generally, pupil size increased after the oddballs. The dilation response was most apparent when the oddball was a white-noise burst embedded in 1-kHz tone bursts as the standards, and was weak when the oddball-standard pair was 1-kHz tone and white noise, or was 2-kHz tone and 1-kHz tone. This preference to the noise oddball was relatively robust to the manipulation of voluntary attention to (or away from) the auditory stimuli. In another experiment, human participants were presented with several kinds of sound bursts, including samples of everyday sounds (e.g., phone calling, animal vocalization, human crying) as well as abstract tones and noises, in isolation of context. The pupil size tended to increase after the presentation of the burst. The dilation size was correlated with a subjective rating on salience. In sum, the present study suggests that the pupil dilation can be used as a physiological marker for certain aspects of auditory salience.

PACS no. 43.64.Ri, 43.66.Lj

1. Introduction

In complex acoustical environments, some sounds are more salient (or attention-capturing) than others. Our current project aims to establish objective measures of the salience of auditory event or object to the listener. The locus coeruleus (LC)-norepinephrine system has been argued to play a key role in attention mechanisms, and the LC neurons activities are reflected to pupil size [1]. Thus, we expect that the auditory salience can be observable as a pupil dilation response (PDR). The PDRs to assumed-to-be salient stimuli have been reported in barn owl [2, 3] and in humans [4-6]. However, systematic understanding is lacking on factors that might contribute to auditory-related human PDRs.

This paper is a preliminary report on experiments to examine the sensitivities of PDRs to factors that are considered to contribute to determining the salience of an auditory event. In experiments 1 – 3, PDRs were measured for low probability “oddball” sounds presented intermittently in a sequence of repeating “standard” sounds. Experiment 1 compared PDRs to 2-kHz- tone and white-noise oddballs presented in 1-kHz-tone standards. Experiment 2 evaluated the effect of oddball probability and of the reversal of the oddball-standard relationship. Experiment 3 examined the effect of voluntary attention towards or away from the auditory stimuli. Experiment 4 was concerned with the salience inherent to a particular sound segment. We measured PDRs to various samples of sound segments presented in isolation of context, and compared the PDRs with subjective rating of the salience.

(c) European Acoustics Association

2. Methods

2.1 Participants

A total of twenty-eight people (aged from 21-47, median: 34 years old, 12 males) with normal or correct-to-normal vision and audition were paid to participate in the current study. Subsets of 7-10 people participated in individual experiments. All participants were naïve about the purpose of the current study. The experimental protocols were approved by NTT Communication Science

Laboratories Ethical Committee, and all participants gave informed consent before the experiment.

2.2 Stimuli and apparatus

Visual stimuli were generated by a personal computer and presented on an 18.1-inch monitor with a resolution of 1024 × 768 pixels. All visual stimuli were presented at the center of the monitor against a gray background. Four types of visual stimuli were used: fixation point, course-grating Gabor patch, fine-grating Gabor patch, and random-dot noise disk. The fixation point was a small gray dot (0.25 × 0.25°) and used for all the experiments. The other visual stimuli were used in experiment 3 only and sized as 5° × 5° and with a matched mean luminance. The Gabor patches were generated by superimposing a Gaussian and a sine wave function with a vertical orientation. The frequency of the course-grating Gabor patch and the fine-grating Gabor patch were 1 and 2 cycles per degree, respectively. The random-dot noise disk was generated by superimposing a Gaussian function and a black-and-white noise squared patch with the intensity in one pixel base.

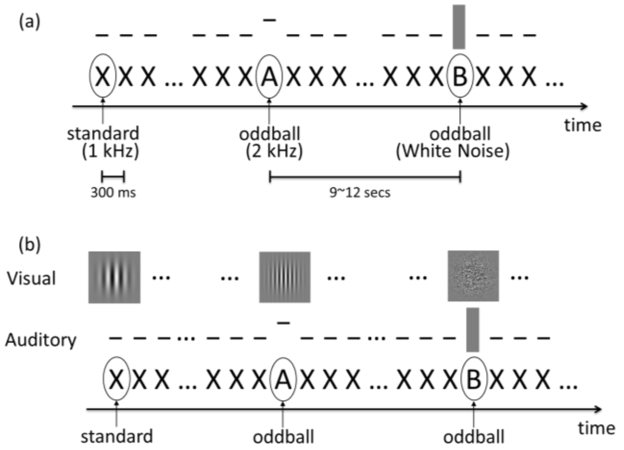

Figure 1. Schematic illustration of stimulus sequence. (a) Stimulus sequence used in experiment 1. A sequence consisted of repeated presentation of tone or noise bursts with 300-ms intervals. The standard stimulus was a 1-kHz tone burst. Infrequent oddball stimuli were either a 2-kHz tone burst or a white noise burst. (b) Stimulus sequence for experiment 3. Auditory stimuli were essentially the same as those in experiment 1. In experiment 3, a visual, coarse-grating gabor patch was presented synchronously to the auditory standard stimulus. A fine-grating gabor patch or random-dot disc was presented with an auditory oddball stimuli. The combination of the visual and oddball stimuli were randomized for individual presentation.

Auditory stimuli were synthesized by a personal computer (sampling rate: 44100 Hz), digital-to-analog converted, amplified, and presented diotically through headphones.

In Experiments 1-3, three types of auditory stimuli were used: bursts of 1-kHz tone, 2-kHz tone (not used in Experiment 2), and white noise. All the stimuli had a duration of 50 ms (including 50-ms raised cosine ramps) and an A-weighted sound pressure level of 65 dB.

An auditory sequence was presented in which oddballs were presented against repeated standard sounds (see illustration in Fig. 1). In experiment 1, the standard sound was the 1-kHz tone burst (1-kHz standard), and the oddball was the 2-kHz tone burst or the white noise burst (2-kHz oddball, and noise oddball, respectively). The two types of oddballs were presented in a randomly assigned order within the same block of auditory sequence. Each sound, including standard and oddball sounds, was presented in 300-ms inter-stimulus-intervals (ISI). Oddballs were separated by 9-12-sec intervals. Each type of oddball was presented 40 times; therefore there were 80 trials of oddballs in total.

In experiment 2, standard/oddball type (tone, noise) and probability (1%, 2%, 10%) were manipulated. In the tone oddball condition, 1-kHz oddballs and noise standards were used. In the noise oddball condition, in contrast, noise oddballs and 1-kHz standards were used. In the 1% probability condition, each sound was presented in 300-ms ISI and oddballs were separated by 18-20 sec. In the 2% probability condition, each sound was presented in 300-ms ISI and oddballs were separated by 9-12 sec. In the 10% probability condition, each sound was presented in 1500-ms ISI and oddballs were separated by 9-15 sec. There were 60 trials of oddballs in each block.

In experiment 3, the auditory sequence was the same as in experiment 1, and a visual sequence was presented simultaneously with the auditory sequence. All the visual stimuli were presented for 50 ms and synchronized with the auditory stimuli. The course-grating Gabor patch was always presented simultaneously with the 1-kHz standard. The fine-grating Gabor patch and the random-dot noise disk were used as visual oddballs and presented simultaneously with the auditory oddballs. The content of the visual oddballs and the auditory oddballs were unrelated, namely, the combination of the visual and auditory oddballs was randomized for individual presentation.

Stimuli for Experiment 4 were 10 sound samples, consisting of artificial abstract sounds (1-kHz *tone*, white *noise*, linear tone-*chirp* from 0.1 to 8 kHz, and *beep* with a fundamental frequency of 500 Hz), and samples of natural sounds extracted from a CD archive (*dog* barking, *bird* calling, *phone* call, *laughter*, and child *crying*) or recorded in our laboratory (*scratch* sound on a black board). Each sound had a duration of 500 ms and an A-weighted sound pressure level of 65 dB. In pupil size measurement, each block of measurements consisted of 10 repetitions of the 10 stimuli in a random order, with ISI of 10 sec.

2.3 Procedure

2.3.1 Pupil size measurement and analysis

All the participants were given written and oral explanation of the nature of the experiment and the pupil response recording. Participants sat in front of the monitor at a viewing distance of 80 cm in a dark chamber, with their head fixed on a chinrest. They were given the auditory oddball sequence diotically through the headphone while their pupil responses were recorded.

In experiments 1, 2 and 4, participants were asked to fixate at the central fixation point throughout the experiment. They were not involved in any task but just listened to the auditory sequence. In experiment 3, participants were asked to perform a discrimination task as soon and accurate as possible, by pressing corresponding buttons in a response box. In the attend-audition condition, they were asked to discriminate whether the auditory oddball was the 2-kHz tone or the noise, and ignore the visual stimuli. In the attend-vision condition, they were asked to discriminate whether the visual oddball was the fine-grating Gabor patch or the random-dot noise disk, and ignore the auditory stimuli.

Pupil diameter was measured binocularly by an infrared eye-tracker camera of the EyeLink system (SR Research) with sampling rate of 1000 Hz. The camera sat just below the monitor. Participants were instructed to blink naturally during the experiment.

Data during the blinks were treated as missing data. Only one eye of data was submitted to analysis, which was the one with fewer missing data due to blinks and recording errors. In each condition, certain trials of standard sounds were randomly chosen to serve as the baseline of pupil response to any kind of auditory stimulation. There were 40 trials of standard sounds chosen in each condition of experiments 1 and 3, and 60 trials chosen in each condition of experiment 2, to match the trial number as each type of the oddball sounds. The interval between chosen standard sounds was controlled to be longer than 9 secs, to avoid accumulated effect across trials within the analysis window (-1~7 secs to stimulus onset). For each trial, raw data was normalized by subtracting the mean and dividing by the standard deviation derived from the data during the period of 1 sec before the stimulus onset. The PDR functions shown in Figs 2-5 represent the median of the normalized pupil size across trials and participants.

In experiment 1-3, differences of PDR magnitude among conditions were evaluated in terms of mean (normalized) pupil size over the 1-3-sec. period after the stimulus onset. Statistical tests on the mean pupil size was conducted by the bootstrap method with a criterion value of $p = 0.05$ to reject the null hypothesis.

2.3.2 Subjective evaluation of salience

With the same participants and 10 sound samples in experiment 4, we estimated subjective magnitudes of the salience of the sounds, using a pair-wise comparison method. In each trial, the participant was required to choose more salient sound (i.e., the sound that would stand out) from a sound pair chosen from the 10 sounds. A sound was presented when the participant pressed a corresponding button on the computer screen. The participant was allowed to listen to the sounds as many as he or she wanted, and made a judgment after listening to each sound of the pair at least once. All possible pairs of the 10 sound samples were tested, and one trial was run for each pair. Thus, there were a total of 45 trials. Obtained scores pooled across participants were used to derive the "salience scale" by the Thurston method [7], on which the 10 samples were mapped.

3. Results

3.1 Context-related salience (experiments 1-3)

3.1.1 Experiment 1

Median of normalized pupil size data of all trials obtained in experiment 1 is plotted as a function of time related to stimulus onset in Fig. 2. For both oddballs, the pupil size started to increase within ~0.5 sec. after the stimulus onset. For the noise oddball, marked increase of pupil size lasted ~4 sec. For the 2-kHz oddball, on the other hand, the pupil size recovered after the peak at ~0.5 sec. The mean of the median pupil size over the 1-3-sec. period was statistically different between the noise and 2-kHz oddball conditions.

3.1.2 Experiment 2

Experiment 2 examined the dependence of the PDR on stimulus context and probability. We tested the effect of altering the oddball-standard relationship by comparing two conditions in which 1-kHz oddballs were presented in noise standards, and noise oddballs were presented in 1-kHz standards. We also varied the probability of oddballs (1%, 2% and 10%).

The pattern of the results differed between the tone- and noise-oddball conditions. In the noise oddball condition (Fig. 3a), significant PDRs to the oddball were observed, and there was no marked differences among oddball probabilities. In the tone oddball condition, a statistically significant oddball-related PDR was observed only for the lowest oddball probability (1%).

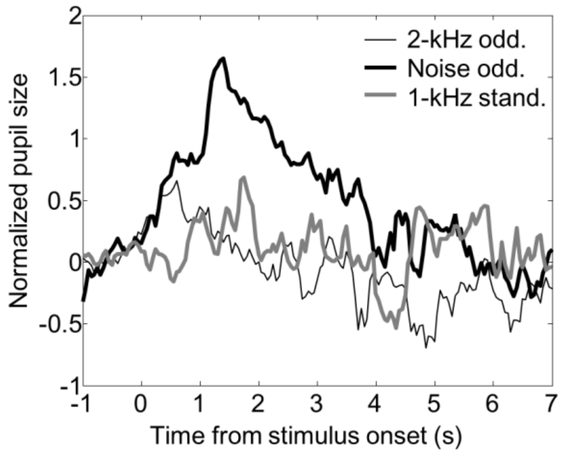

Figure 2. Normalized pupil size as a function of time after the oddball onset obtained in experiment 1. Each function represents the median PDRs across participants and trials for the condition.

3.1.3 Experiment 3

Experiment 3 attempted to evaluate the effect of voluntary attention. Oddball-related PDRs were generally much larger than those found in experiments 1 and 2 (Compare scales of Figs 2-4). These large PDRs are possibly due to the fact that the participants were involved in psychophysical tasks in experiment 3, in which the alertness level of the participants was high, and decision-making and motor-response are required. Main interests of this experiment were not in such absolute size of the responses, but in the relative values across conditions.

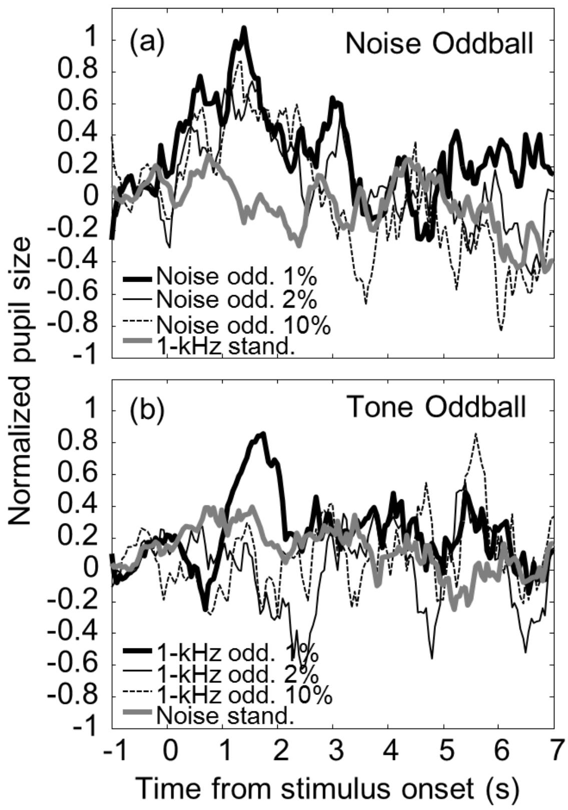

Figure 3. Normalized pupil size as a function of time after the oddball onset obtained in experiment 2. (a) Results when the standard was a 1-kHz tone burst, and the oddball was a white noise burst. (b) Results when the standard was a white burst, and the oddball was a 1-kHz tone burst.

In both attentional states, PDRs tended to be greater to the noise oddball than for the 2-kHz oddball, although the difference between the two oddball conditions in the attend-auditory ($p = 0.06$) only reached marginal significance to reject the null hypothesis.

3.2 Stimulus inherent salience (experiment 4)

The aim of experiment 4 was to explore sensitivities of the PDR to auditory salience due to the inherent characteristics of sound samples. We measured PDRs to various kinds of sound samples, and examined the extent to which the PDR correlates with subjective rating of the salience of the sounds.

Figure 5a shows PDRs to the 10 sound samples, plotted as a function of post stimulus time. One can see that the PDR functions vary across the stimuli.

The same 7 participants participated in an independent psychophysical experiment, in which subjective salience of the sound samples were evaluated with the pairwise comparison method. The Thurston method [7] was used to derive a "salience scale", onto which the sound samples

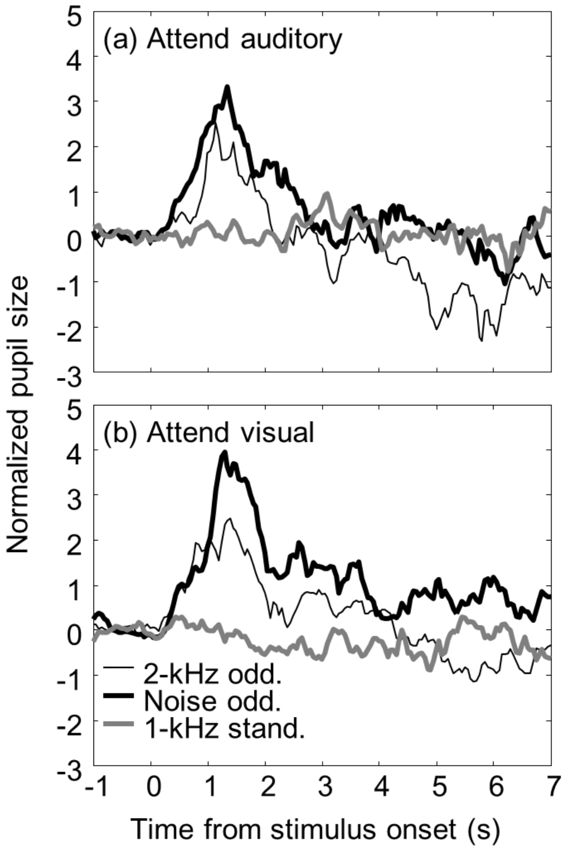

Figure 4. Normalized pupil size as a function of time after the oddball onset obtained in experiment 3. (a) Results when the participants performed the auditory task. (b) Results when participants performed the visual task.The functions were based on only trials in which the participants made correct responses to assure that the participants directed their attentions to specified modality.

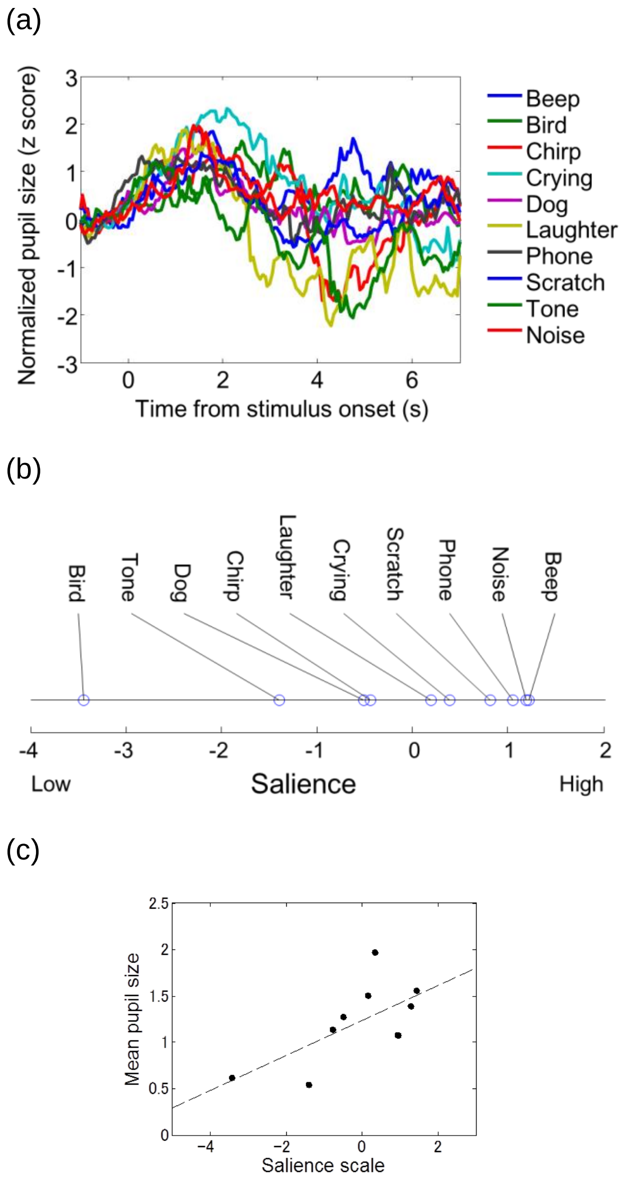

Figure 5. (a) Median of the normalized pupil diameter derived from all trials of all participants as a function of time relative to stimulus onset. (b) Salience scale derived by the psychophysical experiment. The sound samples are mapped on to the scale. (c) Correlation between mean of the median pupil size (1-2 sec. after simulus onset) and salience rating scale (Pearson $r = 0.66$, $p < 0.05$).

were mapped (Fig. 5b). We found a positive correlation between the pupil size and the salience scale, which was statistically significant ($r = 0.66$, $p < 0.05$, Fig. 5c). The results indicate that the PDR reflects the salience of sound segments, or other sound properties that are correlated with the auditory salience.

4. Discussion

The present experiments demonstrated that, as in other species (barn owl: Bala & Takahashi, 2000; monkey: Wang et al., 2014), auditory oddball stimuli elicit PDRs in humans (experiments 1-3). PDRs were observed for sounds presented in isolation, and the response magnitude correlated with subjective rating of the salience of the sounds (experiment 4). The preference for noises over tones observed in experiments 1-3 with an oddball paradigm (described later) was consistent with subjective salience rating of the tone and noise measured in an independent experimental paradigm (experiment 4; compare the salience scores for noise and tone in Fig. 4b). These observations support a hypothesis that the PDR reflect the salience of auditory events or objects.

There was a strong asymmetry in the response magnitude between the noise and tone oddballs. Noise oddballs generally elicited stronger PDRs than 2-kHz oddballs against a background of repetitive 1-kHz tone presentation (experiments 1 and 3). PDRs to the oddballs were more reliable when the oddball was noise in tone standards than when the tone oddballs were presented in noise standards (experiment 2). This asymmetry may reflect relatively higher salience of noise than that of tone. Nevertheless, it is also possible that this asymmetry in PDR reflected the preference of a low-level sensory mechanism that is sensitive to simple acoustical characteristics of the stimuli, such as stimulus bandwidth and loudness. Animal physiological studies suggest an involvement of the superior colliculus in PDR (e.g., [3, 8, 9]). The noise preference of PDR may reflect activities of auditory neurons in the SC, which are known to respond more robustly to broadband than narrowband stimuli (e.g., [10, 11]). Difference in loudness also may explain the asymmetry. Although we equalized the stimuli in terms of A-weighted sound pressure level, the noise could have a larger loudness.

Modulation of voluntary attention did not markedly influence the noise preference of PDR, although the difference in PDR magnitude under the attend-auditory condition only reached the marginal significant in the statistical test (experiment 3). The result suggests that PDR is an orienting reflex to auditory salience, which is relatively independent of voluntary attention.

The results that the probability of oddball presentation had little effect of the PDR magnitude (experiment 2) were against our expectation under the hypothesis of PDR reflecting auditory salience. We expected that the PDR would increase with decreasing probability, associated with increasing novelty-related salience. It is possible that in the current experimental setting, the sensitivity of PDRs to novelty-related salience, if existed, was not observed, dominated by sensitivities to stimulus-related salience.

Acknowledgement

This work was supported by Japan MIC SCOPE, 121803022.

References

[1] G. Aston-Jones, J. D. Cohen: An integrative theory of locus coeruleus-norepinephrine function: adaptive gain and optimal performance. Annu Rev Neurosci 28, 403–450, 2005.

[2] A. D. S. Bala, T. T. Takahashi: Pupillary dilation response as an indicator of auditory discrimination in the barn owl. J Comp Physiol A 186, 425–434, 2000.

[3] S. Netser, S. Ohayon, Y. Gutfreund: Multiple manifestations of microstimulation in the optic tectum: eye movements, pupil dilations, and sensory priming. J Neurophysiol 104, 108–118, 2010.

[4] J. Qiyuan, F. Richer, B. L. Wagoner, J. Beatty: The pupil and stimulus probability. Psychophysiol, 22, 530–534, 1985.

[5] R. M. Stelmack, D. A. Siddle: Pupillary dilation as an index of the orienting reflex. Psychophysiol, 19, 706–708, 1982.

[6] G. Z. Steiner,R. J. Barry: Pupillary responses and event-related potentials as indices of the orienting reflex. Psychophysiol, 48, 1648–1655, 2011.

[7] L. L. Thurstone: A law of comparative judgment, Psychol Rev, vol. 34, p. 273, 1927.

[8] C. A. Wang S. E. Boehnke, L. Itti, D. P. Munoz: Transient pupil response is modulated by contrast-based saliency. J Neurosci 34, 408–417, 2014.

[9] C. A.Wang, S. E. Boehnke SE, B. J. White, D. P. Munoz: Microstimulation of the monkey superior colliculus induces pupil dilation without evoking saccades. J Neurosci 32, 3629–3636, 2012.

[10] A. J. King, S. Carlile: Responses of neurons in the ferret superior colliculus to the spatial location of tonal stimuli, Hear Res, 81, 137-149, 1994.

[11] L. Z. Wise, D. Irvine: Auditory response properties of neurons in deep layers of cat superior colliculus, J Neurophysiol, 49, 674-685, 1983.