Biologically Inspired Cognitive Architectures 24 (2018) 87–97

Contents lists available at ScienceDirect

Biologically Inspired Cognitive Architectures

journal homepage: www.elsevier.com/locate/bica

Research article

Computational modeling of environment deviant sound detection based on human auditory cognitive mechanism

Yu Su[a,b,*], Jingyu Wang[a], Ke Zhang[a], Kurosh Madani[b]

[a] Northwestern Polytechnical University (NPU), 127 Youyi Xilu, Xi'an 710072, Shaanxi, PR China
[b] Université Paris-Est, Signals, Images, and Intelligent Systems Laboratory (LISSI/EA 3956), University Paris-Est Creteil, Senart-FB Institute of Technology, 36-37 rue Charpak, 77127 Lieusaint, France

ARTICLE INFO

Keywords:
Acoustic attention simulation
Auditory deviancy
Sound deviancy detection
Shannon entropy
Sample entropy

ABSTRACT

One of the important aspects of our acoustic perceptual skills is auditory deviancy detection. This acoustic mechanism allowed human beings to percept the novel stimulate while regardless of the processes engaged in the focal task. A similar ability, imitating the above-mentioned human's auditory awareness mechanism will greatly improve the efficiency of artificial perception in complex environment. In this paper, we propose a computational model to mimic such human auditory perception mechanism to verge on this goal. The prosed model consists of two modules: temporal deviancy detection and frequency saliency detection. Combining the information issued from each of the aforementioned modules, the proposed model generates the image indicator that identify the deviant salient-sound which elicit auditory attention shift. The sounds recorded from the real environment have been used for verifying the advantages of the proposed model. The results show that the proposed model is able to point out the deviant salient-sound in a mixture sound clip and shows an acceptable robustness and accuracy. Furthermore, a more comprehensive experiment is performed and illustrate that the proposed model could effectively simulate human auditory attention mechanism.

Introduction

Eyes and ears are the two major sensory organs of human perception system, they cope with myriad stimulus of the surrounding environment almost all day. Receiving this tremendous amounts of stimulus, our brains is capable to extract the pertinent information constructing our cognitive awareness about the environment in which we evolve. Research works relating cognitive psychology (Frintrop, Rome, & Christensen, 2010) have shown that the human's saliency-based selective attention mechanism greatly contributes to human's perception of surroundings and in his actions' efficiency regarding his interactions with the environment. In other words, this cognitive perceptual mechanism acts as a foremost process in construction of our effective awareness about the surrounding environment, helping us to focus on the objects, sounds or events which is conspicuous to us and to reject those (objects, sounds or events) which appear as background noise regarding the target we deal with at a given time. Furthermore, it is a common experience that during we focus on one salient event, our attention can be involuntarily engaged by visual or acoustic changes occurring unexpectedly in the environment (Escera, Alho, Winkler, & Näätänen, 1998; Schröger, 1996). This attention shift phenomenon of our cognitive perceptual mechanism could also be introduced as deviancy detection.

Deviancy detection aims at recognizing situations in which unusual events occur. It seems like the definition of deviancy detection is similar to saliency detection, in fact, they are different in nature. The main purpose of saliency detection is to identify those features in a scene are conspicuous based on their context and are salient, and could attract attention. While the main purpose of deviancy detection is to identify the unusual or deviant events when we focused on the objects or events which attract attention at first. For example, when we listen to music at home, someone knocked on the door, the attention will shift from the music to the knocking. The deviancy detection mechanism could be regarded as a supplement to saliency detection, a bottom-up selection mechanism made up of both helping us to perceive the environment more precisely.

Compared with visual signals, sound signals will enable mankind to be aware of and avoid danger beforehand or when human vision is not available in certain environment. After decades of development, the visual perception of artificial intelligence has been greatly improved. If we can build a comprehensive auditory perception ability for artificial intelligence, then it will greatly improve their perception performance

* Corresponding author at: Northwestern Polytechnical University (NPU), 127 Youyi Xilu, Xi'an 710072, Shaanxi, PR China.
E-mail addresses: suyunwpu@163.com (Y. Su), jywang@nwpu.edu.cn (J. Wang), zhangke@nwpu.edu.cn (K. Zhang), madani@u-pec.fr (K. Madani).

https://doi.org/10.1016/j.bica.2018.04.002
Received 12 March 2018; Accepted 9 April 2018
2212-683X/

of the complex environment. However, most of the auditory selective attention mechanism-based technology are mainly focused on sound saliency detection (Duangudom & Anderson, 2007; Kalinli & Narayanan, 2007; Kaya & Elhilali, 2012; Kayser, Petkov, Lippert, & Logothetis, 2005; Wang, Zhang, Madani, & Sabourin, 2015). The research on acoustic deviancy detection is still in the theoretical research stage (Menon & Uddin, 2010; Vachon, Labonté, & Marsh, 2017).

In this context, the focus of this paper is modeling the sound deviancy detection mechanism of human beings. A background on the recent literature on the topic is provided in the following section.

Related works

Anomalous sounds which could cause auditory attention shift possess two characteristics: (1) It is salient in the entire sound clip; (2) It is deviant relative to the salient sounds that has been detected or focused before. Therefore, the purpose of our goal is to detect the sounds have both the above two properties and irrespective the prominent sound that has been detected before.

For the auditory saliency detection part, since the research results of neuropsychology (Itti, Koch, & Niebur, 1998) proved that visual and auditory perception channels have perceptual correlations in high-level perceptual processing, then it's reasonable to expect that the perception of auditory saliency could be convert into the perception of saliency of the visual channel. This result provides a theoretical foundation and a new approach to realize computational models of ASD. Based on this, several ASD models have been proposed for salient sound detection. An auditory saliency map (ASM) which based on Itti's work is first proposed by Kayser.

Kayser initially proposed an auditory saliency map for salient sound detection. Experiment results showed that this model could mimics several basic properties of human auditory perception mechanism. (Kalinli & Narayanan, 2007) proposed an innovative ASM in for improving the performance of Kayser's model, the new model added the orientation and pitch as new sound features. There are two hypotheses proposed in (Shamma, 2001) for the coding of pitch in human auditory system: temporal hypothesis and spectral hypothesis. Here temporal hypothesis has been chosen for the pitch features extraction and then project it to the frequency axis of spectrogram to generate the feature map of pitch. The result shows that the detection accuracy of prominent syllable and word can reach 75.9% and 78.1%, respectively. (Duangudom & Anderson, 2007) proposed another ASD model in which the time-frequency receiver domain model and adaptive suppression were used to provide the final auditory saliency map. The model presented in this paper is basically the same as Kayser's auditory saliency map, but the features extracted from sound spectrogram and the method of yielding final ASD is different.

These models are generally considered as three classical auditory saliency detection models. Based on the classical models, many researches have proposed several new ASD models during these years. In (Kaya & Elhilali, 2012), an auditory saliency map which treat the input signals as a one-dimensional temporal input was presented. In (Kim, Lin, Walther, Hasegawa-Johnson, & Huang, 2014), a saliency detection model based on the classification result was presented. (Tsuchida & Cottrell, 2012) and (Schauerte & Stiefelhagen, 2013) introduced their novel auditory saliency map based on the theory of statistics to predict the saliency in soundscapes. The recordings in meeting rooms or simple natural sounds to verify the efficiency of these models. Since the environment sounds do not contain a definite pitch and its instability characteristics, we can predict that when these models are applied in environment salient sounds detection tasks, their performance could be seriously affected. In (Wang et al., 2015) proposed a composite system that combined parallel paths including: temporal analysis, spectral analysis and the image salience model. It is reported that this model provided better robustness to saliency detection especially in real noisy soundscapes then conventional methods.

For the deviant sound detection part, the current study mainly aims to reveal the response and mechanism of auditory cortex to deviant sound through electro-encephalograph (EEG) and mismatch negativity (MMN) auditory evoked potential. It is reported in (Escera et al., 1998) that small changes in the acoustic environment capture attention involuntarily by activating the stimulus-change detector mechanism reflected in the mismatch negativity (MMN). Through the study of anterior insula (AI) and considered it as a hub of a "salience network", a network model is presented in (Menon & Uddin, 2010) for better understand brain mechanisms in important environmental stimuli detection tasks. Two parallel but separate lines of research on auditory novelty detection is presented in (Escera & Malmierca, 2014) and indicated that auditory novelty system should be organized in a hierarchical manner. In (Escera, Leung, & Grimm, 2014), after review the evidence of three kinds of human brain response to deviant sounds, the author concluded that deviance detection is a basic principle of the functional organization of the auditory system. By assessing the sensitivity of Middle-Latency Responses components to deviant probability manipulations, the study of (López-Caballero, Zarnowiec, & Escera, 2016) further characterize the auditory hierarchy of novelty responses. In (Kaya & Elhilali, 2013), a biologically motivated model is proposed to building a computational model of MMN based on Kalman filters. This model is tested by finding the deviant onset times of simple oddball paradigms and simple sound patterns. The study of the relationship between human pupillary dilation response (PDR) and deviant auditory stimuli (Liao, Yoneya, Kidani, Kashino, & Furukawa, 2016) showed that a salient event which is deviant from the background attracts attention and reflected in the PDR. The experiment results presented in (Vachon et al., 2017) demonstrate that the deviation effect reflects a general form of auditory distraction as interference took place both within and across domains and regardless of the processes engaged in the focal task.

In general, deviancy detection is a key characteristic of the auditory system that allows pre-attentive discrimination of incoming stimuli irrespective the ongoing constant stimulation. Hence, providing artificial intelligence with such auditory mechanism will effectively enhance its perceptual performance in real environment.

In this paper, we proposed a computational model to capture the deviant salient-sound in real environment which mimics human auditory attention shifting mechanism. This approach is based on the detection of deviant salient-sounds in the temporal domain combined with the frequency domain saliency detection and present the detected deviant sound in the image domain at last. The model first obtains the local salient sounds in the time domain through a combined feature of Gammatone Filterbank Cepstral Coefficient (GFCC). Then, an entropy-based analysis method is applied to find the sound with deviancy which elicit the acoustic attention shift. Second, the sound saliency in the frequency domain is derived from the Power Spectral Density (PSD) based frequency saliency detection method and been considered as frequency deviancy of sounds. Afterwards, in the opponent color space, the gammatone spectrogram blue-yellow channel information is calculated as the indicator to present the deviant salient-sound which lead to the auditory attention shift.

Method

Modern artificial machines still could not perceive its surrounding environment as intelligent as human does, because the perception process is mostly based on the visual sensory information rather than auditory. Therefore, it is essential to provide human-like approach for machine to form the artificial computational awareness ability while the feature of human acoustic characteristic should be considered in order to mimic the human hearing property. Furthermore, the environmental sound signals are varying in both temporal domain and frequency domain while the auditory deviancy detection has some similarity with auditory saliency detection to some extent. Hence, we will