Biologically inspired assessment of noticeability of sound events in context

Karlo Filipan, Bert De Coensel, Sarah Verhulst, Dick Botteldooren
Department of Information Technology, Ghent University, Ghent, Belgium

Summary
Annoyance caused by environmental noise intruding the private dwelling and perception of the sonic environment in public spaces share a critical dependence on the detection of salient sound events. During everyday activities, the probability of noticing a sound in the complex sonic environment is proportional to how much this sound stands out of its context. Based on a thorough review of human auditory processing, scene analysis and attention, we propose a computational model that allows to identify salient sounds in a complex environment. The tonotopic model possesses a unique capability to trace amplitude modulations and phase sweeps, which are features that the human auditory system is highly sensitive to. The model is validated by exploring its response to sound environments with known annoying characteristics such as short rise times and impulses and results are contrasted against Zwicker loudness.

PACS no. 43.50.Rq, 43.50.Yw, 43.66.Lj

1. Introduction

Soundscape research has shown that sounds which are noticed influence the perception of soundscape [1]. While walking through an urban environment, people generally pay little attention to details in the audiovisual environment when not asked to do so [2]. Most environmental sounds may therefore remain unnoticed and hence would not contribute to the cognitive appraisal of the sonic environment. However, some sounds have intrinsic characteristics that separate them from their background. The measure of separation is sound saliency—the degree of how much a sound stands out in the sonic environment.

The two most important saliency traits are sensory saliency, which is determined by the enhanced sensitivity or tuning of the human hearing system for specific sound features, and semantic saliency, which requires recognition of the sound and incongruency within the environment [3]. Sensory saliency has been investigated by explicitly identifying features that increase the behavioral response or by resemblance of the spectrogram with visual saliency [4]. By associating sensory saliency to the tuning of the human auditory system, brain imaging techniques can also serve as a starting point for creating computational sound saliency models. In [5] it was shown that topographically localized regions of the brain respond to specific spectrotemporal sound modulations, i.e. amplitude and frequency modulation ripples.

Several models for evaluating saliency based on modulations have been proposed throughout the years [6, 7, 8]. Building on that knowledge, a model for sensory saliency that accounts for amplitude and frequency modulations on top of a tonotopically organized representation of sounds in the auditory system [9] was improved and evaluated. The model was created with soundscape research in mind—it utilizes techniques that enable a constant stream of input while the simplification of the computationally expensive calculations enables it to run in real time on a smaller device. Therefore, this biologically inspired model would enable analysis of large amounts of data available in soundscape studies thus bridging the gap between highly complex auditory neuroscience models and simple indicators used in soundscape research.

2. Computational models for noticeability evaluation

The stages of the computational model for auditory saliency are presented on Figure 1. The acoustic input is fed to an auditory periphery model [10] that simulates how the sound is represented up to the level of the auditory brainstem. The model output at each tonotopic regio is used as an input to the model of higher band processing: auditory cortex and sensory activation.

The auditory cortex stage is modeled using spectrotemporal modulation representations [11]. In this

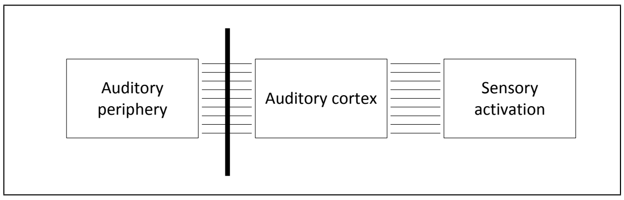

Figure 1. Stages of the computational model for auditory saliency—output of auditory periphery is used as an input to the simulation of higher brain processing [11].

stage, an input consisting of several frequency band signals is passed through a series of resonator filters with their resonance on different amplitude modulations (AM). Afterwards, the signal on each frequency band is delayed using the buffers with the length corresponding to the frequency modulation (FM) on the exact frequency band. The next step includes overlapping summation across several frequency bands, thus reducing the number of output bands. Finally, to remove the rippling effect in the output, the maximum is taken across buffers with the length related to the amplitude modulations. Therefore, at each time step, the output consists of three dimensions: AM, FM and (a reduced number of) frequency bands. Sensory activation is modeled by excitation and inhibition integration with different rise and fall time constants [9, 11]. Finally, the overall saliency is calculated by summing over all outputs of the sensory activation stage.

Several models for the auditory periphery have been published previously [12, 13, 14]. Nevertheless, in this paper we compared the two most distinct options: fast implementation with Gammatone filterbank and complex simulation of the physical processes that take place in the auditory system [10]. On the one hand, a fast implementation is necessary for being able to use the model with large datasets usually available in soundscape studies. On the other hand, complex simulation of the auditory system is desired when the model is used for simulating and evaluating more detailed characteristics of human hearing.

In general, for determining noticeability of sound events, soundscape studies apply simplified models using commonly established psychoacoustic indicators [15]. Therefore, in this study, two additional models based on such indicators were included: Zwicker loudness according to the ISO 532-1 standard [16] and the energy-equivalent continuous A-weighted sound pressure level (L_{Aeq}). We evaluate our model against those standard approaches to see if the biologically inspired models perform better.

3. Evaluation of environmental sounds

The models were compared using two groups of sound signals. A first batch was created by combining background traffic noise with a 1000 Hz pure tone beep at five different levels (”Peak in noise”). It should

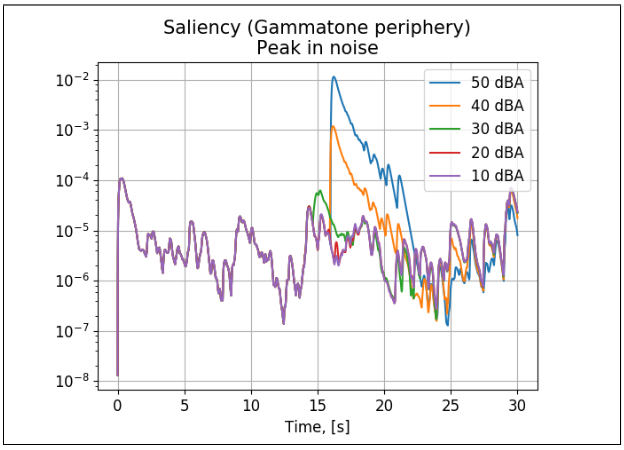

Figure 2. Response of the computational saliency model based on spectrotemporal modulations with fast Gammatone periphery on different levels of beep in traffic noise.

be noted that the sound was created with randomized time of the beep start, to remove the influence of traffic noise characteristics in the output detection. In turn, such signals would determine at which input level the signal-to-noise ratio (SNR) of the beep to background noise in the input would enable a clear detectability in the output. Furthermore, a second batch of sounds was created by varying the rise time (”Slope change”) of a particular segment of industrial noise, which would demonstrate if and how much the rise in the input influences the magnitude of the output [17].

3.1. Changing the level of a beep in traffic noise

The response of the saliency model based on spectrotemporal modulations with Gammatone auditory periphery on different levels of the 1000 Hz beep is displayed on Figure 2. As it can be seen, the peaks corresponding to the onset of the beep are clearly detected up to 20 dB(A) of the lowest level of the beep. On the other hand, for the same auditory saliency implementation but with complex ear model periphery only the highest two peaks are detected (Figure 3) which is due to the masking implemented in the ear periphery.

Furthermore, although the same parameters were used for the higher brain processing stage, the saliency output with the ear periphery exhibits a faster decline from the peak (i.e. shorter fall time). This trait stems from the inhibition already included in the response of the auditory nerve fiber. At the same time, the loudness model (Figure 4) as well as the amplitude of the equivalent level (Figure 5) exhibit no decrease of the output for the duration of the tone. Finally, for the 30 dB(A) beep, the amplitude calculated from equivalent continuous A-weighted level demonstrates better detectability of the onset peak than the loudness model output.

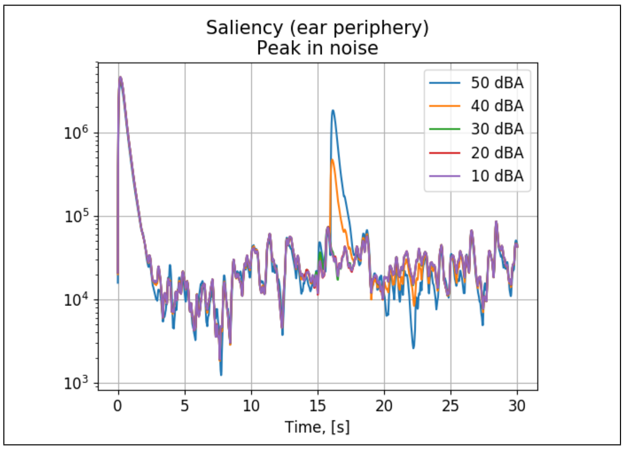

Figure 3. Response of the computational saliency model based on spectrotemporal modulations with complex ear model periphery on different levels of beep in traffic noise.

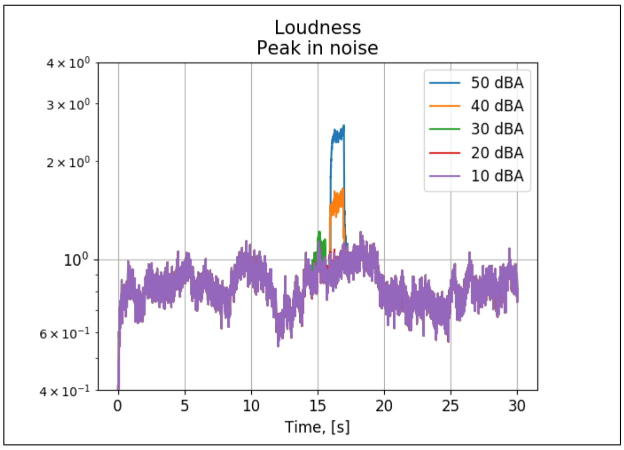

Figure 4. Zwicker loudness on different levels of beep in traffic noise.

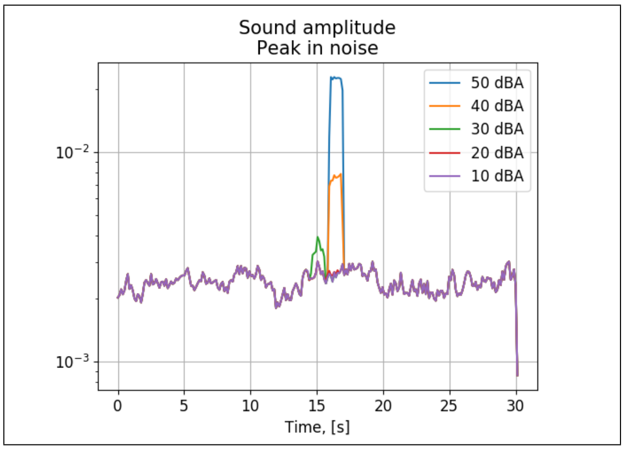

Figure 5. Equivalent continuous A-weighted sound amplitude on different levels of beep in traffic noise.

3.2. Changing onset slope of the industrial noise

The output of the four models in relation to rise time was checked using the industrial noise with varying

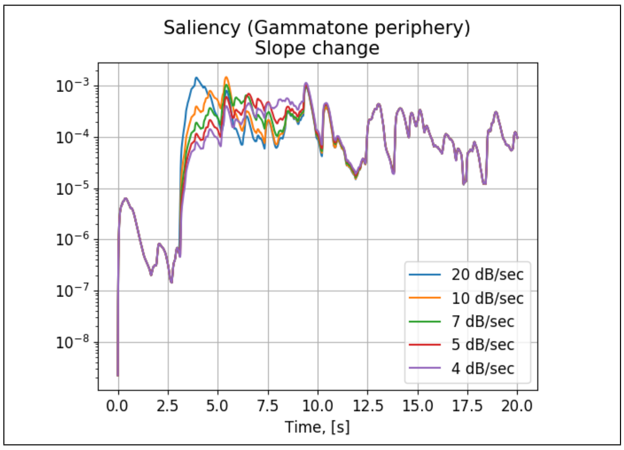

Figure 6. Response of the computational saliency model based on spectrotemporal modulations with fast Gammatone periphery on different rise times of the industrial noise.

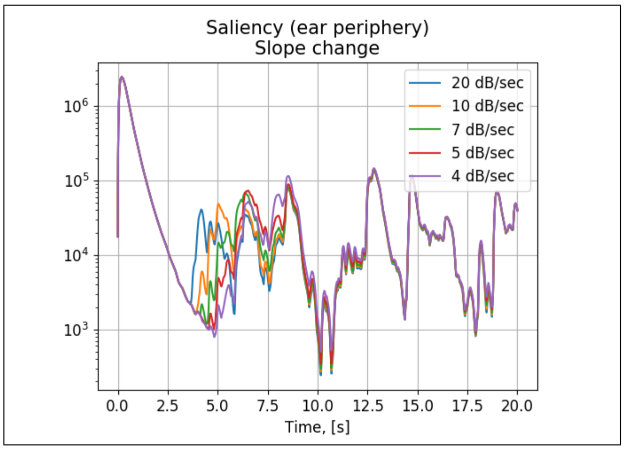

Figure 7. Response of the computational saliency model based on spectrotemporal modulations with complex ear model periphery on different rise times of the industrial noise.

onset slopes (from the fastest rise of 20 dB/s down to the slowest 4 dB/s). The saliency model with Gammatone periphery (Figure 6) displays an increase in saliency magnitude for the steepest onset, therefore showing the influence of the higher onsets on the saliency output.

Furthermore, for the saliency model with ear periphery (Figure 7) the change was detected in the latency difference with relation to different rise times corresponding to the fact that the ear reacts faster for steeper rises.

Finally, the loudness and sound amplitude (Figures 8 and 9) were found not to be dependent on the slope, but their outputs were instead following the rise up to a steady signal proportionally to the onset.

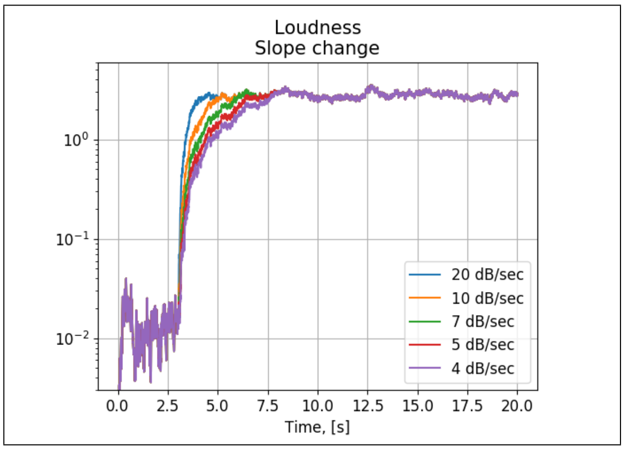

Figure 8. Zwicker loudness on different rise times of the industrial noise.

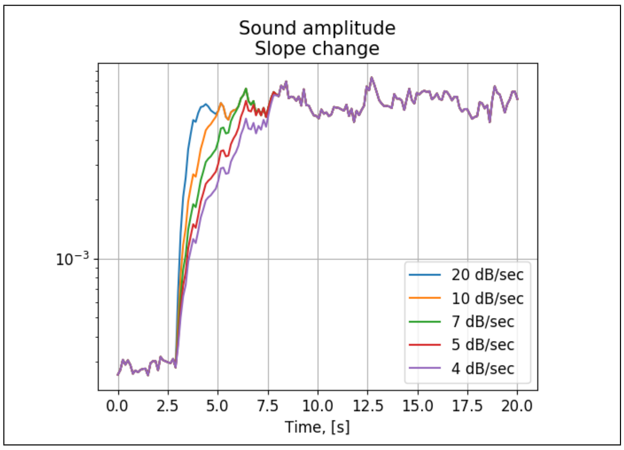

Figure 9. Equivalent continuous A-weighted sound amplitude on different rise times of the industrial noise.

4. Conclusions

In this paper we presented the assessment of noticeability of representative sound events that might occur in environmental sound context. The comparison included four models appropriate for evaluation of sound events in soundscape research: saliency based on spectrotemporal modulations with Gammatone periphery, saliency based on spectrotemporal modulations with ear periphery, Zwicker loudness and finally energy-equivalent continuous A-weighted sound pressure level.

It was shown that the biologically inspired auditory saliency model based on spectrotemporal modulations with Gammatone auditory periphery had clear detection of the beep signal inside the traffic noise even for low signal-to-noise ratios. Moreover, the same model had a higher response for the fastest rise time in comparison to other models. Consequently, the shown examples indicate that the computational model of auditory saliency based on spectrotemporal modulations adds valuable information for evaluation of noticeable events occurring in a sonic environment.

To further investigate the appropriateness of the proposed model, future studies will include evaluation of the stimuli used in auditory neuroscience experiments as well as the artificially created sound signals relevant for soundscape studies. Finally, the model will be validated with experimental data obtained through a continuous evaluation of sound saliency during walking trips through urban environments.

Acknowledgement

This research has received funding from the Research Foundation Flanders (FWO-Vlaanderen) under grant n° G0D5215N, ERC Runner-up project MAESTRO.

References

[1] J. Kang, F. Aletta, T. T. Gjestland, L. A. Brown, D. Botteldooren, B. Schulte-Fortkamp, P. Lercher, I. van Kamp, K. Genuit, A. Fiebig, J. L. B. Coelho, L. Maffei, L. Lavia: Ten questions on the soundscapes of the built environment. Building and Environment 108 (2016) 284-294.

[2] J. A. Droll, M. P. Eckstein: Gaze control and memory for objects while walking in a real world environment. Visual Cognition 17, no. 6-7 (2009) 1159-1184.

[3] A. Zlatintsi, P. Koutras, G. Evangelopoulos, N. Malandrakis, N. Efthymiou, K. Pastra, A. Potamianos, and P. Maragos: COGNIMUSE: a multimodal video database annotated with saliency, events, semantics and emotion with application to summarization. EURASIP Journal on Image and Video Processing 2017, no. 1 (2017) 24.

[4] C. Kayser, C. I. Petkov, M. Lippert, N. K. Logothetis: Mechanisms for allocating auditory attention: an auditory saliency map. Current Biology 15, no. 21 (2005) 1943-1947.

[5] M. Schönwiesner, R. J. Zatorre: Spectro-temporal modulation transfer function of single voxels in the human auditory cortex measured with high-resolution fMRI. Proceedings of the National Academy of Sciences 106, no. 34 (2009) 14611-14616.

[6] T. Chi, R. Powen, S. A. Shamma: Multiresolution spectrotemporal analysis of complex sounds. The Journal of the Acoustical Society of America 118, no. 2 (2005) 887-906.

[7] E. M. Kaya, M. Elhilali: Investigating bottom-up auditory attention. Frontiers in human neuroscience 8 (2014) 12.

[8] E. M. Kaya, M. Elhilali: Modelling auditory attention. Phil. Trans. R. Soc. B 372, no. 1714 (2017) 10.

[9] K. Filipan, A. Bockstael, B. De Coensel, M. Schönwiesner, D. Botteldooren: A novel auditory saliency prediction model based on spectrotemporal modulations. Proc. 22nd International Congress on Acoustics (2016).

[10] S. Verhulst, A. Altoè, V. Vasilkov: Computational modeling of the human auditory periphery: auditory-nerve responses, evoked potentials and hearing loss. Hearing Research (2017) 21.

[11] K. Filipan, B. De Coensel, S. Verhulst, D. Botteldooren: A biologically inspired auditory saliency model for urban soundscape analysis and design. In preparation.

[12] V. Hohmann: Frequency analysis and synthesis using a Gammatone filterbank. Acta Acustica united with Acustica 88, no. 3 (2002) 433-442.

[13] F. M. Rønne, T. Dau, J. Harte, C. Elberling: Modeling auditory evoked brainstem responses to transient stimuli. The Journal of the Acoustical Society of America 131, no. 5 (2012) 3903-3913.

[14] M. S. A. Zilany, I. C. Bruce, L. H. Carney: Updated parameters and expanded simulation options for a model of the auditory periphery. The Journal of the Acoustical Society of America 135, no. 1 (2014) 283-286.

[15] F. Aletta, J. Kang, and Ö. Axelsson: Soundscape descriptors and a conceptual framework for developing predictive soundscape models. Landscape and Urban Planning 149 (2016) 65-74.

[16] ISO 532-1:2017 Acoustics – Methods for calculating loudness – Part 1: Zwicker method. International Organization for Standardization (2017) 58.

[17] T. H. Pedersen: Audibility of impulsive sounds in environmental noise. Proc. 29th International Congress on Noise Control Engineering. 4158-4164.