

University of Szeged
Doctoral School of Computer Science

Hardware-Software Co-development for Audio and Video Data Acquisition and Analysis

Summary of the Doctoral Theses

György Kalmár

Supervisor:

László G. Nyúl, PhD



Szeged
2020

1 Introduction

Data science is the algorithmic identification of patterns in data. However, the data may be diverse and pattern recognition applications include many different fields such as signal processing, image analysis, and speech recognition, to name a few. A key component of these analysis pipelines is the data they are trained and tested on, and its quality fundamentally limits the accuracy of the results. Important factors are the number of data points, their appropriate distribution, and noisiness. As good the data is, as accurate the methods can get.

The low-level, hardware-related changes are undesirable in data acquisition and analysis systems as they usually imply the modification of the software as well. However, if hardware-caused drawbacks influence data quality, even the high-level data analysis methods are affected. Sophisticated algorithms may reduce these drawbacks slightly, but spending processing capacity to correct the anomalies is a waste of time and energy — two such quantities that are limited, for example, in embedded systems. Many times applications reach a stage where further software optimizations offer minimal improvements compared to the cost and time required to implement them. Noisy, low-quality data can be enhanced, but with restrictions. In these cases, even simple hardware changes or extensions may lead to improved data quality and new analysis directions. I applied this consideration in my PhD work.

The PhD thesis presents three data analysis applications that include embedded audio classification systems and image segmentation methods. A common approach connects them that is with hardware-oriented modifications and software co-development, the high-level tasks became feasible, simpler, or more accurate.

The dissertation consists of three major parts. In Chapter 2, an animal-borne gunshot detector system is presented that was improved by a novel wake-up mechanism. Another acoustic event detection related topic is investigated in Chapter 3 that studies the loudspeakers' sound recording capabilities, which option became feasible with a simple hardware extension. In Chapter 4, a video processing application is detailed that automated and improved the pupillometry of rats to support schizophrenia-related medical research.

2 Animal-Borne Anti-Poaching System

Poaching is listed among the top five drivers of biodiversity loss. Interventions to reduce it typically follow classic law enforcement approaches, however, poaching of the wildlife tends to occur in remote areas with low human densities, where detection is difficult. Also, poaching of large, high-value species is militarized and supported by global crime syndicates. As such, local wildlife agents are operationally overwhelmed, not only in terms of law enforcement equipment but often due to the limited capacity to monitor widely distributed animals. The development of technologies designed to overcome the challenges of remote wildlife protection is needed. Nowadays, a promising direction is the utilization of animal-borne sensors, particularly GPS-equipped collars, which are used to enhance real-time wildlife protection.

In the second chapter of the thesis, an animal-borne acoustic gunshot detector was introduced that extended the functionality of widely-used GPS tracking collars. With the

fusion of the two systems, gunshot alerts can be raised in real-time coupled with location data. The main challenges were the multi-year lifetime requirement, the preservation of the recorded ballistic shockwave sound quality, and the minimization of the false positive gunshot detection rate.

2.1 A novel wake-up mechanism

Acoustic gunshot detection is a pattern recognition problem that requires the constant recording and processing of environmental sounds. There are two acoustic events associated with firing a typical rifle and both of them can be picked up by a microphone. The muzzle blast is the result of the propellant of the ammunition exploding inside the barrel of the gun. The other event is called the ballistics shockwave, and it is caused by the bullet traveling faster than the speed of sound. This shockwave is a unique acoustic phenomenon, its shape in the time-domain resembles a capital *N* with sub-microsecond rise time and a total signal length of a few hundred microseconds. These characteristics make it an ideal target signal for an animal-borne gunshot detector.

The combination of the ultra-low power consumption and the recording of the shockwaves with good quality is essential in wearable gunshot sensing. To satisfy both requirements, a novel wake-up mechanism was proposed. It is based on an acoustic delay line structure illustrated in Figure 1. This solution uses two types of microphones: a contact and a traditional electret microphone. The contact microphone, or pickup, is a transducer that converts the vibration of the surface it is mounted on to voltage by utilizing the piezoelectric effect. In our case, this microphone is attached to the internal side of the detector module enclosure. The second, traditional, microphone is placed at the end of a 3.5 cm-long tube. This tube serves as a waveguide and soundproofing for the incoming sound waves.

The idea behind this structure is to wake up the data acquisition system from deep sleep mode when the acoustic waves reach the enclosure wall and delay the sound waves by the tube to ensure the required amount of time for the acquisition system to prepare for data collection. It is possible since the speed of sound is negligible compared to the speed of light and the voltage generated by the contact microphone travels at the latter. With this wake-up mechanism, the power consumption was reduced by 88% compared to the traditional solutions.

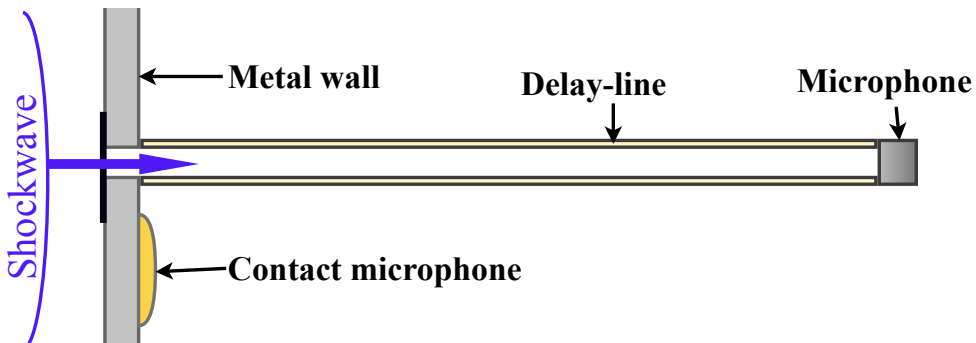


Figure 1: *The proposed wake-up delay line structure.*

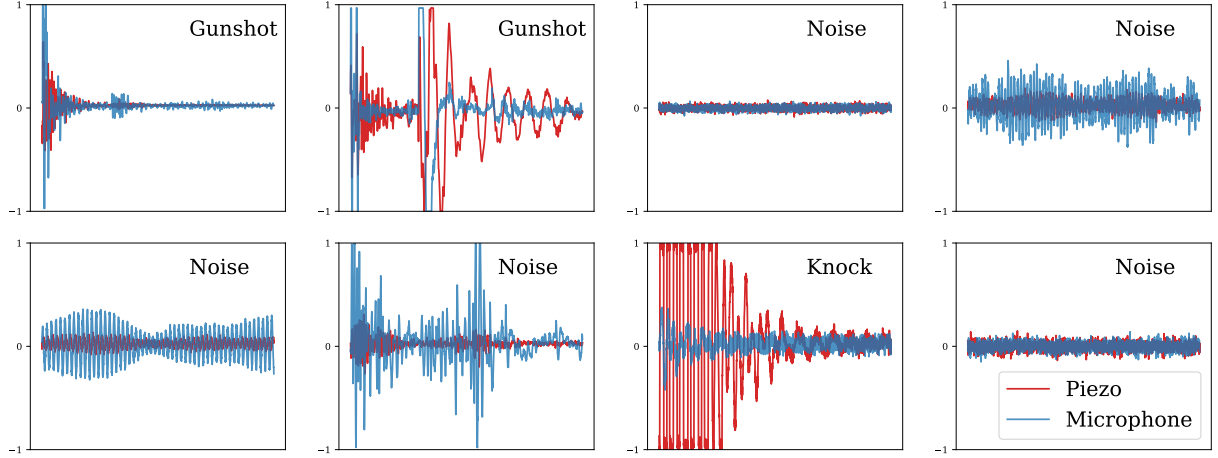


Figure 2: Example recordings acquired through the acoustic delay line. This novel structure employs two microphones separated by a short tube.

2.2 Application in GPS tracking collars

The proposed gunshot detector has a multi-year lifetime using a single D-cell battery, which makes it ideal for wearable applications. The hardware and software of this gunshot detector module were optimized to fit into widely-used GPS tracking collars used for elephants. Given the strength of elephants, the mechanical protection of an acoustic sensor is challenging. The protective material needs to be strong and thick to survive the required lifetime of the sensor. The used steel box, thick nylon top, and the resin filling offer reliable protection but also reduce the acoustic signal quality inside the enclosure. To compensate these unwanted effects, a small hole was drilled into the metal wall to enable unattenuated sound propagation into the box, and this hole was covered by an acoustic waterproof vent.

The size of the existing protective box was already fixed. We attached our gunshot detector sensor board to the wall by a compact microphone and sensor board holder unit, which I designed to be 3D printed from a semi-soft rubbery material. The holder’s most important feature is the embedded acoustic waveguide. A 3.5 cm-long, 3 mm-wide tube, without any sharp turns or edges, is meandering inside the holder, connecting the hole in the metal wall with the microphone. This curved tube design reduced the size of the holder to about half of the tube length. The other side of the holder unit presses the contact microphone to the metal surface of the enclosure.

The first prototype of the proposed system was deployed on a wild elephant in Kenya. To evaluate the system, controlled real-world experiments were also carried out. A dataset was collected that contained environmental sounds, mechanical impact noises, and real gunshots. A set of example recordings are shown in Figure 2.

2.3 Gunshot detection

After an acoustic event has happened, the system starts recording it within 100 μ s, and a fast, nearly real-time decision is needed because the risk of sensor damage is very high as poachers may try to destroy the device. Therefore, processing time must be limited, mandating the use of simple algorithms, and the resource-constrained embedded platform also points in the same direction.

The proposed gunshot detector has three stages (see Figure 3). The first stage runs in real-time, and its main function is to filter out false wake-up events. The second stage implements cross-domain filtering and only runs offline. Its main objective is to filter out acoustic events caused by mechanical impacts on the box. The two microphones with different characteristics simplify this task. When a mechanical impact happens, the contact microphone’s signal gets clipped while the electret microphone signal amplitudes remain small. In contrast, when an acoustic wave reaches the device, only a small portion of the energy is converted to vibration resulting in small amplitudes in the piezo microphone signal while the electret microphone signal is pronounced. The third stage of the detector implements the most complex analysis that mainly relies on shockwave detection. It is based on the unique N -wave shape and symmetries of the shockwave. Muzzle blast detection and further filtering methods were also implemented to provide additional confirmation for shockwave detections, but they cannot generate alarms separately. In Figure 3, a set of examples can be seen; the recording of a knock, an animal sound, and a gunshot. It also illustrates the basic structure and behavior of the detector, and the propagation of the recordings through the stages.

The gunshot detection algorithm achieved good results on the collected acoustic event dataset as all the gunshots were successfully detected, and no false positive alarms were generated. Data-driven methods were also briefly mentioned in the thesis, and a randomized architecture-search algorithm was developed that generated and trained 1D and 2D convolutional neural networks for gunshot detection, and compared them.

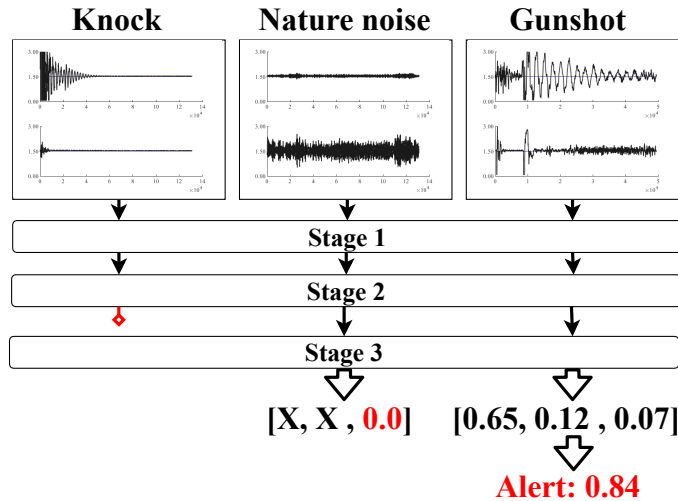


Figure 3: The structure and the behavior of the detector with three example recordings that propagate through the stages. In these recordings on the top, the upper signals correspond to the contact microphone, the lower ones to the electret microphone.

3 Reverse Mode Speakers

The loudspeaker is an electroacoustic transducer that converts an electrical signal into sound. The most widely used type is the dynamic loudspeaker that produces sound by forcing a coil with an attached diaphragm to move rapidly back and forth. However, it is well-known that speakers can record sound as well, and this "microphone" mode can be referred to as *reverse mode*. This reversed behavior is similar to the working principle of dynamic microphones. The incoming sound waves exert force on the surface of the diaphragm, which starts vibrating, inferring the oscillation of the coil in a magnetic field. As magnetic field fluctuations occur through the coil, electromotive force is being generated, i.e. a voltage difference builds up between the coil's two terminals. This varying voltage represents the incoming sound in the electrical domain.

The proposed idea in the third chapter of the thesis was the utilization of reverse mode speakers in acoustic event detection applications. The hardware extension that offers this extra functionality is minimal and can be implemented by a simple embedded device. This device provides the original, radiating mode operation but extra, microphone-like capabilities also become feasible. For example, such extended speakers could be employed in security applications, where suspicious acoustic event detection is required.

3.1 Theoretical analysis

I started the investigation of the reverse mode with theoretical modeling. An equivalent mechanical circuit was formed and the reverse mode transfer function was derived, which can be calculated for loudspeakers from their published parameters. The equivalent circuit and its simplified form is presented in Figure 4. The analysis of this transfer function showed that reverse mode speakers have a frequency region with enhanced sensitivity

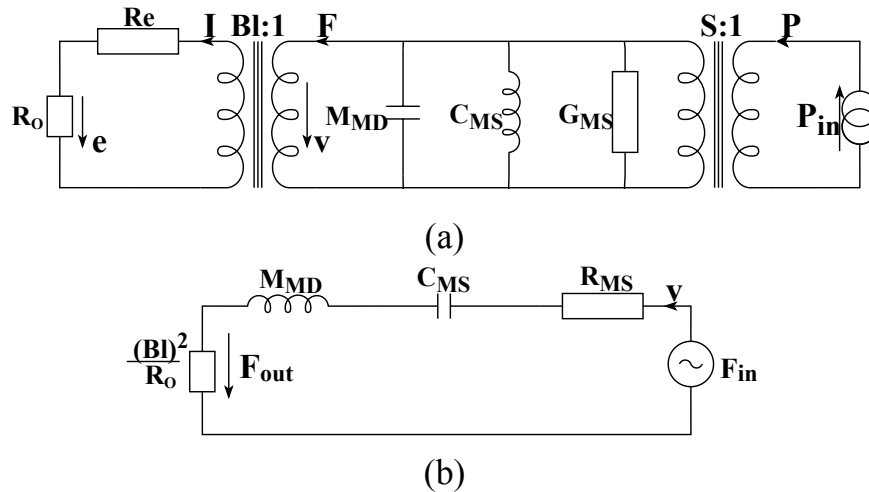


Figure 4: Equivalent circuits of the reverse mode. In (a), virtual transformers are used to represent the force and electrical signal generation steps and the driving source is placed into the acoustic domain. The electrical output signal can be measured on R_o . In (b), the simplified mechanical equivalent circuit is presented after the elimination of the virtual transformers. The elements are transformed into the mechanical domain by using mechanical impedance.

around their resonance frequency. Real measurements demonstrated that this sensitivity is sufficient for acoustic event recording with proper quality.

After it was shown that their capabilities are acceptable for audio recording, I examined the reverse mode speakers in event detection scenarios and carried out simulation-based experiments by using their transfer functions. These experiments simulated reverse mode speaker responses to microphone-recorded events and ran classification algorithms on them to show the effects of the information loss introduced by the low and non-linear reverse mode sensitivity.

The simulations converted audio files recorded by microphones into forms as they would have been recorded by a particular reverse mode speaker. After transforming an urban sound audio dataset, I could highlight the strengths and weaknesses of the reverse mode speakers through the analysis of the performance of neural networks trained on the transformed data for audio event classification. From the classification results, it was concluded that reverse mode speakers could be used for event detection. However, the type and nature of these events are limited. For example, reliable speech recognition could hardly be achieved from reasonable distances because of the low sound pressure levels and low reverse mode sensitivity. At the same time, loud, impulsive events like gunshots, explosions, screaming, etc. could be detected with sufficient accuracy. These observations can be explained based on the conclusions obtained from the theoretical analysis. The reverse mode sensitivity is low outside the resonance frequency zone, therefore, the information content of loud events is more likely to be preserved. Furthermore, impulsive events have wider spectra, thus at least the frequency components close to the highly sensitive resonance frequency are recorded with a good quality.

3.2 Utilization perspectives

The utilization of reverse mode speakers could be interesting in areas, where already deployed speakers are available and with minimal hardware changes new applications would become achievable. For example, if a building contains multiple speakers that are interconnected and are driven by the same central driving source, the whole area could be protected by deploying a single device that is listening on the driving line. Similar setups can be found in hospitals, stations or schools, where short statements are occasionally announced through the speakers, but between these active periods, the inactive speakers could be used for event detection. In Figure 5, two such scenarios and recorded events are presented. Another potential utilization direction is smartphones that have at least two loudspeakers and one microphone. In these highly integrated devices, the introduction of new hardware parts may result in huge costs, but the already included speakers' utilization could be maximized by extending the capabilities to record their reverse mode signals. With this modification, smartphones would have three audio input channels, thus complex, sound-based localization applications would become feasible.

To present these potential directions, I designed and implemented an audio event detector device based on the reverse mode functionality, and demonstrated its use by a simple, data-driven clap detector.

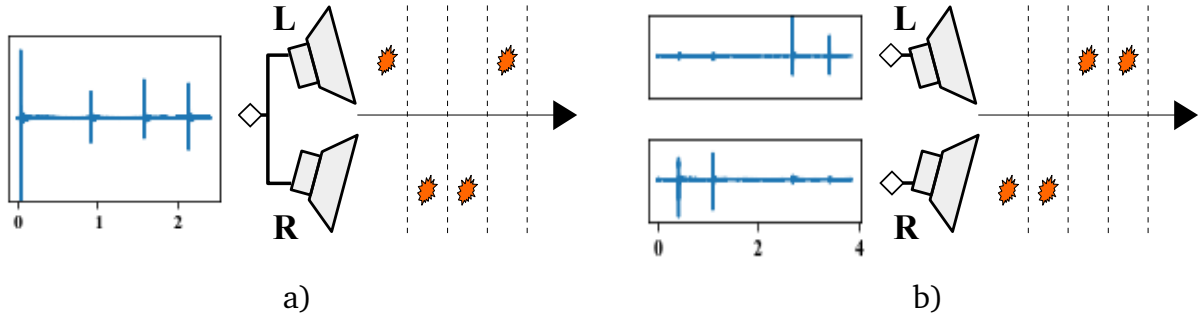


Figure 5: Possible configurations of the reverse mode speakers in event detection setups. In (a), the speakers are connected, thus both generate their reverse mode signals on the same driving line, and in (b), the speakers have their own dedicated lines. During the experiments, four claps were produced with different relative locations to the speakers.

3.3 Active reverse mode

In the previously mentioned event detection scenarios, the speakers were inactive, which means that their driving sources were inactive during the recording phases and the idea was to utilize these idle periods. However, the work analyzed also the active reverse mode scenario, where acoustic event detection is required while the speakers are being actively driven in parallel with the acoustic events. These situations are typical in public spaces (stores, cafes, restaurants, shopping centers, etc.), where low volume music is played constantly from distributedly deployed loudspeakers. These places may become the targets of violence against the public or terror attacks.

During the active reverse mode, the speaker is actively producing sound while external acoustic waves are also reaching the cone. In this case, the sum of the driving and reverse mode electric signals is present on the driving line, which behavior is explained by the superposition of the driving and reverse mode mechanical forces that act at the same time on the diaphragm. In Chapter 3 of the thesis, I presented theoretical modeling and analysis, and some experiments to examine this interesting behavior.

4 Automated Pupillometry

Pupillometry is a long-known method that is used to objectively characterize the pupil's response to light stimuli. Traditional pupillometry experiments record the pupillary light reflex (PLR), i.e. the contraction of pupil in response to light, and then, the pupil diameter is measured in each video frame to produce the pupillogram. Measuring the diameter manually in each frame with an annotation software is time consuming and non-reproducible, especially in cases, when the number of videos is high. Computer algorithms, however, may provide faster, more reliable, and reproducible solutions. In Chapter 4 of the thesis, the automation of a pupillometry application was presented.

The related medical research aims to reveal objectively detectable effects of schizophrenia on the nervous system in a rat model through the comparison of the PLR curves of healthy and test animals with schizophrenia-like symptoms. Developing animal models for any psychiatric diseases, such as schizophrenia, is essential to understand the disorder.

4.1 Pupillometry with classical algorithms

The pupillometry animal experiments took place in a dark room, therefore, the rats' pupils were dilated. The room was illuminated by infrared light, which is invisible for the rats but can be recorded by an infrared camera. During the recorded part of the experiments, rats were held by hand on a desk while a short visible light impulse was emitted into their eyes, which induced the PLR reaction. Before the experiment, the camera was located close to the eye to record the response.

The videos had quality drawbacks. The breathing and slight movements of rats caused motion artifacts and significant blur. Furthermore, the examined rats were albinos with red eyes that reduced the contrast between the pupil and the iris regions. Low lighting conditions forced a higher ISO level value, which implied noisier images. The reflections of the illuminating infrared LEDs obscured a big part of the pupil boundary, while other overlying entities such as the whiskers made many times the pupil nearly undetectable. To overcome these challenges and accurately detect the pupil and automatically measure its parameters, a novel ray propagation based image processing method with an energy attenuation model was developed. This algorithm could find the fine contrast changes at the boundary of the pupil and tolerated noise, reflections, occlusions and blurred images.

The proposed method used notions and ideas taken from the physics of ray propagation. The rays have an initial energy that is gradually absorbed by the medium during the propagation. The amount of energy loss is proportional to the attenuation coefficient of the medium. Based on these principles, the concept is to radiate rays from a point to all directions and use the pixel intensities as the measures of attenuation coefficients. The method traces the rays while they travel through the image and uses the information of energy loss characteristics to learn more about the structure of the surrounding regions. This method produced the estimated boundary points of the pupil, filtered them, carried out robust ellipse-fitting to the retained points, and calculated the pupil diameter. This automated method was evaluated on 20 manually annotated videos and achieved a low, 2% average relative pupil diameter error.

The videos were processed by the diameter measurement software that produced the corresponding pupillograms. To compare the responses of the healthy and test animals, descriptive features from these pupillogram curves were extracted. These features are relevant from the pathophysiological point of view and suitable to emphasize the differences between the groups regarding the autonomic nervous system activities. To support this feature extraction phase, I developed an automated method that produced 40 features from a pupillogram. Also my contribution was the introduction of new dynamics- and smoothness-related features. The extracted features served as the basis for data analysis and the obtained results were published in a medical journal that revealed impaired pupillary control related to schizophrenia in the investigated rat model [2].

4.2 Improved pupillometry

The automated pupillometry led to significant medical results in the field of schizophrenia-related research. However, the experimentation setup set video quality and robustness limitations and many experiments had to be repeated and many videos could not be processed. To improve the robustness of the experiments, the measurement setup was redesigned. An IR LED-ring was attached around the camera lens, which configuration

placed the camera optical axis and the illuminating IR LEDs (nearly) on the same axis. With this modification, the camera could capture the light reflected back from the retina, which induced the bright pupil effect. Here, the pupil region operates as a small "window" that lets the incoming and reflected light through, thus it glows bright, while the other structures scatter the incoming light and remain dark in the recorded image. This effect enhanced the signal-to-noise ratio and improved the detectability of the pupil. Furthermore, this configuration allowed the camera to be placed farther from the eye, thus tiny animal movements did not affect the recording quality.

The videos recorded by the new setup had a completely different nature, therefore, the previously detailed measurement method could not be used. However, the improved video quality reduced the complexity of pupil segmentation and a reasonable amount of data was enough to train a data-driven method. Neural networks offer a modern way to solve a segmentation problem, especially fully-convolutional networks are frequently used in similar problems. For the pupil segmentation task, a popular U-shaped model architecture was utilized. This model was trained on our publicly available manually annotated dataset. On the test images, the diameter predictor achieved 96% median accuracy and the processing time of the video frames reduced significantly.

The comparison between the original and the revised measurement setups and some processed video frames can be observed in Figure 6.

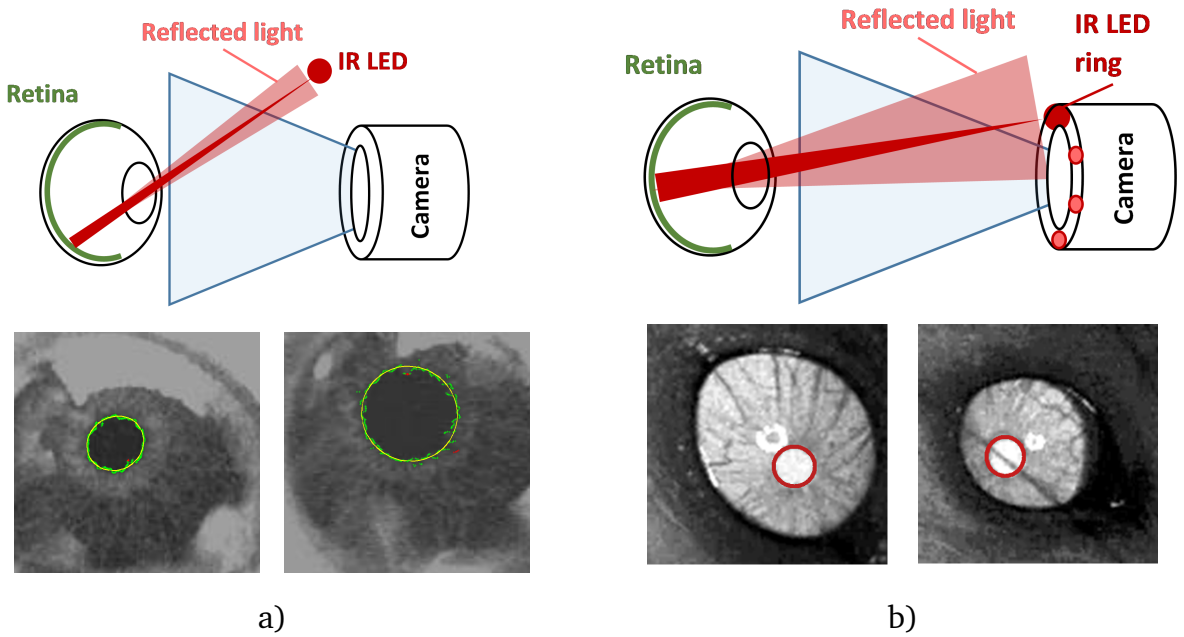


Figure 6: Comparison between the previously used (a) and the revised (b) pupillometry experimentation setups. The enhanced video quality can be observed in (b). In (a), the pupil regions were segmented by a novel image processing algorithm, and in (b), similar results were produced by a convolutional neural network.

5 Contributions of the thesis

In the **first thesis group**, the contributions are related to an ultra-low-power gunshot detector. Detailed discussion can be found in Chapter 2.

- I/1. I proposed a novel acoustic delay line wake-up mechanism, implemented an experimental hardware, and showed that it can improve the power-consumption efficiency of audio event detectors.
- I/2. I designed and implemented the hardware and software of an embedded gunshot detector module that utilizes the proposed wake-up mechanism and can be integrated into widely-used GPS tracking collars.
- I/3. I developed a novel gunshot detector algorithm that employs the two-domain audio information used for the proposed wake-up mechanism, and evaluated its accuracy and efficiency through real-world experiments.
- I/4. I developed a randomized architecture-search algorithm that generated, trained, and compared 1D and 2D convolutional neural networks that utilize the two-domain audio information for gunshot detection.

In the **second thesis group**, the contributions are related to the theoretical and practical investigations of the microphone mode of loudspeakers (referred to as reverse mode). Detailed discussion can be found in Chapter 3.

- II/1. I proposed the idea of using loudspeakers for audio event detection by employing their reverse mode functionality. I carried out the theoretical modeling and analysis of the reverse mode, and supported it by real experiments.
- II/2. I investigated through simulation experiments the reverse mode speakers in acoustic event detection scenarios.
- II/3. I designed and implemented an audio event detector device based on the reverse mode functionality, and demonstrated its use by a simple, data-driven clap detector.
- II/4. I investigated the loudspeakers' active reverse mode through theoretical modeling and analysis, and some experiments, when the speakers may be used for acoustic event detection while they are actively radiating sound.

In the **third thesis group**, the contributions are related to the automation of pupillometry and related image processing methods. Detailed discussion can be found in Chapter 4.

- III/1. I developed and evaluated a pupil measurement method that is based on an energy attenuation model, implemented an automated feature extractor, and introduced new pupillogram features.
- III/2. I redesigned the previously used pupillometry experimentation setup with a hardware extension on the camera, which enhanced the video quality, and thus supports more robust and more efficient experimentation.
- III/3. I trained a fully-convolutional neural network for pupil segmentation that efficiently processes the videos acquired by the revised experimentation setup.

Table 1 summarizes the relation between the thesis points and the corresponding publications.

Table 1: *Correspondence between the thesis points and my publications.*

Publication	Thesis point										
	I/1	I/2	I/3	I/4	II/1	II/2	II/3	II/4	III/1	III/2	III/3
[1]									•		
[3]										•	•
[4]					•		•	•			
[5]	•	•	•	•							
[6]					•	•					
[7]						•	•				

The author's publications on the subjects of the thesis

Journal publications

- [1] **G. Kalmár**, A. Büki, G. Kékesi, G. Horváth, and L. G. Nyúl. Image Processing-based Automatic Pupillometry on Infrared Videos. *Acta Cybernetica*, 23(2), 599-613, 2017.
- [2] A. Büki, **G. Kalmár**, G. Kékesi, G. Benedek, L. G. Nyúl, and G. Horváth. Impaired pupillary control in “schizophrenia-like” WISKET rats. *Autonomic Neuroscience*, vol. 213, 34-42, 2018.
- [3] **G. Kalmár**, A. Büki, G. Kékesi, G. Horváth, and L. G. Nyúl. Automating, Analyzing and Improving Pupillometry with Machine Learning Algorithms. *Acta Cybernetica*, 24(2), 197-209, 2019.
- [4] **G. Kalmár**. Analysis and Utilization of Reverse Mode Loudspeakers. *IEEE Access*, vol. 8., 66270-66280, 2020.

Full papers in conference proceedings

- [5] **G. Kalmár**, G. Wittemyer, P. Völgyesi, H.B. Rasmussen, M. Maróti, and Á. Lédeczi. Animal-Borne Anti-Poaching System. In *Proceedings of the 17th Annual International Conference on Mobile Systems, Applications, and Services (MobiSys '19)*, Association for Computing Machinery, 91-102, 2019.
- [6] **G. Kalmár**. Investigation of Reverse Mode Loudspeaker Performance in Urban Sound Classification. *27th European Signal Processing Conference (EUSIPCO)*, 1-5, 2019.
- [7] **G. Kalmár**. Smart Speaker: Suspicious Event Detection with Reverse Mode Speakers. *42nd International Conference on Telecommunications and Signal Processing (TSP)*, 509-512, 2019.

Further related publications

- [8] **G. Kalmár**, A. Büki, G. Kékesi, G. Horváth, and L. G. Nyúl. Image processing based automatic pupillometry on infrared videos. In *The 10th Jubilee Conference of PhD Students in Computer Science (CSCS): Volume of extended abstracts.*, 2016.
- [9] **G. Kalmár**, A. Büki, G. Kékesi, G. Horváth, and L. G. Nyúl. Feature extraction and classification for pupillary images of rats. In *The 11th Jubilee Conference of PhD Students in Computer Science (CSCS): Volume of extended abstracts.*, 2018.
- [10] **Kalmár G.**, Büki A., Kékesi G., Nyúl L., and Horváth G.. Pupillametria automatizálása, vizsgálata és javítása gépi tanuló algoritmusokkal. *Képfeldolgozók és Alakfelismerők Társaságának 12. Országos Konferenciája*, 2019.
- [11] **G. Kalmár**, G. Wittemyer, P. Völgyesi, H.B. Rasmussen, M. Maróti, and Á. Lédeczi. Animal-Borne Acoustic Gunshot Detector (poster). In *Proceedings of the 17th Annual International Conference on Mobile Systems, Applications, and Services (MobiSys '19)*, Association for Computing Machinery, 578-579, 2019.
- [12] A. Büki, **G. Kalmár**, G. Kékesi, G. Benedek, L. G. Nyúl, and G. Horváth. Characterization of pupillary response in “schizophrenia-like” (WISKET) rats. *5th FENS Regional Meeting 2017*, 2017.
- [13] Büki A., **Kalmár G.**, Kékesi G., Nyúl L., and Horváth G.. Autonóm idegrendszeri eltérések vizsgálata transzlációs modellben. *A Magyar Élettani Társaság, a Magyar Kísérletes és klinikai Farmakológiai Társaság és a Magyar Mikrocirkulációs és Vaszkuláris Biológiai Társaság közös Vándorgyűlése*, 2017.
- [14] Büki A., **Kalmár G.**, Kékesi G., Nyúl L., and Horváth G.. A pupilla fényreflex nem-beli különbségeinek vizsgálata patkányban. In *Magyar Élettani Társaság 2018. évi Vándorgyűlése : előadás és poszter absztraktok*, 2018.