

Titel der Ausarbeitung

Autor A, Autor B, Autor C

Zusammenfassung—

Abstract—

I. EINFÜHRUNG

In nature the ability to hear or in other words the ability to gain informations about your environment by sound processing is an essential skill for many animals as well as for humans. Whether it is for hunting prey, for communication, or for drawing attention to potential threats, audition can help solving a variety of different tasks.

In today's world, humans use their ears in road traffic most of the time. At the same time, the technological advances in the development of autonomous cars are making significant progress. It is reasonable to assume that an auditory system could improve even more the performance of autonomous car driving. While most humans use their car horns to communicate warnings, sirens are an important tool for police, fire and rescue services that are using them officially to indicate an emergency. An auditory system can improve the decision-making of the autonomous car based on informations e.g. on the road character, the car condition, and the squealing of the tires.

In the field of robots and sound source localization, auditory systems became an important research field.

In 2003, Valin *et al.* showed that a mobile robot can localize different types of sound sources over a range of 3 m with a precision of 3° in real time using an array of 8 microphones (<https://ieeexplore.ieee.org/document/1248813/>). By 2016, they were able to localize and track simultaneous different moving sound sources over a range of 7 m using beamforming and particle filtering (<https://arxiv.org/abs/1602.08139>). Liu *et al.* took a different approach with a biologically inspired spiking neural network for sound localisation (https://link.springer.com/chapter/10.1007/978-3-540-87559-8_41) in 2008. Their experimental results showed that their model could localize a sound source from the azimuth angle, the angle of incidence, -90 to 90 degree. In 2009, Murray *et al.* presented a hybrid architecture using cross-correlation and recurrent neural networks for acoustic tracking in robots (<https://link.springer.com/chapter/10.1007/115210825>). Using only two microphones, their model has shown comparable results with the capabilities of the human auditory cortex with the azimuth localisation differing by an average of $\pm 0.4^\circ$. Murase *et al.* used an array of 8 microphones mounted on a mobile robot in order to track multiple moving speaker. Their two key ideas were to use beamforming to locate the sound sources and to use a set of Kalman filters to track the

non-linear movements of the speaker (https://www.isca-speech.org/archive/archive_papers/interpeech2005/i050249.pdf).

The used filters had different history lengths in order to reduce errors under noisy and echoic environments. As a result, multiple moving speakers could be tracked successfully even when speakers and the mobile robot moved non-linearly. So far, most of those systems have in common that they are built to work in closed or crowded environments to interact with people. Focusing on auditive systems for cars, we find that Fazenda *et al.* demonstrated an acoustic based safety emergency vehicle detection for intelligent transport systems in 2009 (<https://ieeexplore.ieee.org/document/5332788/>). Based on a cross microphone array, they were capable of determining the incoming direction of a siren as a sound source. For their suggested array radius, their methods, which were based on time delay estimation, outperformed those, based on calculating the intensity at the microphone array.

So far, most of the mentioned system focused on sound localization, tracking and separation. Another important aspect for a hearing car is source classification. While the goal for most auditory systems for mobile robots is speech recognition, an autonomous car will more likely be confronted with environmental sounds. Performing source classification with neural networks is an active field of research. In 2012 Shen *et al.* (<https://www.sciencedirect.com/science/article/pii/S1474667016334127>) proposed a system based on pattern recognition using a Gaussian mixture models and Mel-Frequency Cepstral Coefficients features. Their system showed an average accuracy of 91.36% for off-line tests with 8 different sound sources. Further evaluation in on-line tests yield good results as well. Pizsak (<https://ieeexplore.ieee.org/abstract/document/7324337/>) used a convolutional neural network to classify short audio clips of environmental sounds. His model outperformed baseline implementations relying on mel-frequency cepstral coefficients and performed comparable to other state of the art approaches. This was validated with 3 public available data sets for environmental and urban recordings.

Baelde (<https://ieeexplore.ieee.org/document/7952592/>)

A. Requirements and constraints

Several requirements have to be fulfilled to enable the practical use of auditory systems for autonomous car driving. A detailed overview can be found in the thesis by Marko Durkovic (<https://mediatum.ub.tum.de/1096033>). The following requirements are necessary for our auditory system:

- **Robustness towards reverberation:** As sound waves propagate through space they get reflected by surfaces

in their environment. A captured sound under real conditions always means that the recording consists of the original sound source and its reflections. The environment of a car can change constantly. Whether it be driving at urban terrain or countryside, a changing environment will yield different magnitude of reverberation which can become extrem while driving through a tunnel for instance. The auditory system for an autonomous car needs to be robust towards this changing environments without losing too much of its accuracy.

- **Robustness towards noise:** Captured sound has usually to deal with two different kind of noise. One part is sensor noise, also known as self-noise introduced by the microphones. The other part are unwanted sound sources created by the environment. In the environment of an autonomous car, the proportion of sensor noise compared to the sound sources of your environment will be small. Sensor noise often becomes a problem with high pre-amplification whereas we will have to deal with high intensity noise given for instance by the airflow around the microphone array or by engine sounds around you. Therefore we assume that the sensor noise is negligible whereas the produced noise by our environment will be a major factor our auditory system has to deal with.
- **General applicability:** The auditory system of an autonomous car will likely be faced with a lot of different type of sound sources. Therefore the system needs to maintain performance even if multiple sources with different signal characteristics are present.
- **Number of sources:** In a dynamic environment like road traffic it is not possible to determine the number of sound sources that the car will get in simultaneous contact with. Therefore the auditory system needs to be able to process observations where the amount of active sound sources is not known beforehand and possibly greater than one. The lower and upper bound of sources the system shall be able to track depends on the number of the events and their likelihood to occur. Limiting the system to identifying horns and sirens an upper bound of two different active sources can be sufficient. The more information the system is supposed to gather, the more complex it becomes.
- **Number of dimensions:** The position of an active sound source can be described by three parameters relative to your own coordinate system. In order to track sirens or horns, the auditory system should at least be able to estimate the direction of the sound source. [That allows combined with tracking for a 2 dimensional localisation. Even though an estimation of the distance of the source would be desirable.??]
- **Source classification:** For the auditory system of an autonomous car the source classification is a mandatory feature. It has to identify sirens and preferably other desired sources with high accuracy. It needs to be robust for noise depending on the preprocessing of the sound.
- **ROS:**

II. GRUNDLAGEN

A. *Modules of an auditory system*

Subsection text.

B. *HARK*

Subsubsection text.

C. *Sound source localization*

Subsubsection text.

D. *Sound source tracking*

Subsubsection text.

E. *Sound source separation*

Subsubsection text.

F. *Filter*

Subsubsection text.

G. *Integration of HARK in ROS*

Subsubsection text.

III. ZUSAMMENFASSUNG

ANHANG I

OPTIONALER TITEL

Anhang eins.

ANHANG II

Anhang zwei.

DANKSAGUNG

Wenn ihr jemanden danken wollt, der Euch bei der Arbeit besonders unterstützt hat (Korrekturlesen, fachliche Hinweise,...), dann ist hier der dafür vorgesehene Platz.

LITERATURVERZEICHNIS

- [1] H. Kopka and P. W. Daly, *A Guide to L^AT_EX*, 3rd ed. Harlow, England: Addison-Wesley, 1999.
- [2] Deutsche Forschungsgemeinschaft, *Vorschläge zur Sicherung guter wissenschaftlicher Praxis*, Denkschrift, Weinheim: Wiley-VCH, 1998.