Anwendungen akustischer Ereigniserkennung im Automobil
Automatic Acoustic Event Detection in Automotive Applications

Jan Rennies[1], Jens Schröder[1], Danilo Hollosi[1], Marten Wittorf[2], Dr. Volker Grützmacher[2], Stefan Goetze[1]

[1]Fraunhofer IDMT, Hör-, Sprach- und Audiotechnologie, Oldenburg, Deutschland, jan.rennies@idmt.fraunhofer.de
[2]Adam Opel AG, Rüsselsheim, Deutschland

Kurzfassung

Die Verwendung akustischer, mittels Mikrofonen gewonnener Signale wurde in bisherigen Anwendungen im Bereich Automobil abgesehen von Freisprechsystemen oder Spracherkennung zur Mediensteuerung überwiegend vernachlässigt, obwohl akustische Signale gegenüber anderen Modalitäten deutliche Vorteile bieten können. So ist es beispielsweise möglich Signale (z.B. Sirenen von Einsatzfahrzeugen) zu detektieren, die sich (noch) außerhalb des Sichtfeldes befinden und damit einen Zeitvorteil zu erlangen. In dieser Studie wurde ein prototypisches Maschinen-Lern-Verfahren zur automatischen Erkennung von Alarmsirenen im Verkehrslärm entwickelt und systematisch untersucht. Es zeigt sich, dass hohe Erkennungsleistungen selbst bei niedrigen Signal-Rausch-Abständen möglich sind. Da sich die eingesetzten Technologien mit vertretbarem Aufwand auch auf andere Signaltypen (z.B. Bremsenquietschen) übertragen lassen, könnte akustische Ereigniserkennung ein vielseitige Ergänzung zu bestehenden Sensoren im Automobil liefen.

Abstract

Apart from speech recognition and hands-free communication systems, signals acquired acoustically via microphones have been largely neglected in automotive applications so far, although they may offer substantial advantages compared to other modalities. For example, they allow to detect signals emitted from acoustic sources (yet) outside the field of vision and thereby to gain time. In this study, a prototypical machine-learning method to automatically detect alarm sirens in traffic noise was implemented and systematically investigated. The results indicate high detection accuracy even at low signal-to-noise ratios (SNRs). Since the employed technologies can also be transferred to other types of signals (e.g., break squeal) with moderate effort, acoustic event detection might represent a versatile extension of existing sensors in automotive applications.

1 Einleitung

Einsatzfahrzeuge, wie z.B. Polizei, Krankenwagen oder Feuerwehr, spielen in täglichen Verkehrssituationen eine besondere Rolle. Für unbeteiligte Verkehrsteilnehmer können sie im gravierendsten Fall eine Bedrohung darstellen, da sie sich in zeitkritischen Situationen nicht an Geschwindigkeitsbegrenzungen, Verkehrszeichen oder Ampelsignale halten müssen. Um die hierdurch entstehende Gefährdung zu verringern, werden verschiedene Arten von Warnsignalen eingesetzt, u.a. optische Lichtwarnsignale und akustische Alarmsirenen. Während optische Warnsignale durch Verkehrsteilnehmer nur dann wahrgenommen werden können, wenn sich der Signalgeber im direkten Sichtfeld befindet, sind akustische Alarmsignale auch dann wahrnehmbar, wenn sich der Signalgeber dem Verkehrsteilnehmer bspw. von hinten oder durch Häuser verdeckt nähert. Diese effiziente Warnfunktion ist jedoch dann eingeschränkt, wenn die Hörwahrnehmung erschwert ist, z.B. durch das Hören von lauter Musik, lautes Verkehrsgeräusch oder eine mögliche Schwerhörigkeit des Verkehrsteilnehmers [1].

Bisherige Ansätze zur Erkennung von akustischen Alarmsirenen lassen sich im Wesentlichen in zwei Kategorien unterteilen: Regelbasierte Verfahren (z.B. [1, 2, 3, 4, 5]) nutzen einzelne oder mehrere physikalische Eigenschaften der Signale, für die die aktuell vorliegenden Werte gegen definierte Schwellwerte geprüft werden. Bei Überschreiten der Schwellwerte wird angenommen, dass das vorliegende Signal einem Alarm entspricht. Für derartige Verfahren wurden im Allgemeinen geringere Erkennungsraten bzw. höhere Fehlalarmraten berichtet. Im Gegensatz dazu wurden vielversprechende Ergebnisse mit sogenannten Maschinen-Lern-Verfahren erzielt (z.B. [6, 7]). Diese Verfahren lernen die Unterscheidungsregeln zwischen Zielsignal und akustischem Hintergrund automatisch auf Grundlage von Trainingsmaterial und einer Repräsentation durch akustische Merkmale (Engl. Features).

Das Ziel dieser Studie war es das Potenzial automatischer akustischer Ereigniserkennung zur Detektion und Klassifikation von verkehrsrelevanten Szenarien am Beispiel akustischer Alarmsirenen systematisch zu untersuchen, wobei Maschinen-Lern-Verfahren verwendet wurden. Abschnitt 2 beschreibt die dafür eingesetzten Methoden und Signale. Die Ergebnisse werden in Abschnitt 3 beschrieben und in Abschnitt 4 diskutiert.

2 Methoden

2.1 Trainings- und Testsignale

Für diese Studie wurde eine Datenbank verschiedener Zielsignale (Alarmsirenen) und Hintergrundgeräusche (Verkehrslärm) erstellt. Die Zielsignale wurden aus unterschiedlichen Web-Datenbanken sowie aus eigenen Aufnahmen von acht realen Polizeifahrzeugen gewonnen. Da sich die Schallquellen und Empfänger in der Praxis im Allgemeinen relativ zueinander bewegen, wurden die Signale mit einer Simulation des Dopplereffektes für Relativgeschwindigkeiten von -50 bis +50 m/s nachbearbeitet. Die Signale wurden so geschnitten, dass jeweils genau eine Sequenz bestehend aus dem tiefen und hohen Ton der Sirene enthalten war. Diese Abschnitte wurden zu zufällig gewählten Zeiten in ein 5 s langes Ruheintervall platziert. Insgesamt wurden auf diese Weise 378 reine Alarmsignale gewonnen.

Als Hintergrundgeräusche dienten zwei unterschiedliche Verkehrsgeräusche, die ebenfalls in Abschnitte von 5 s geschnitten wurden. Ein Geräusch entsprach typischem Verkehrslärm bei trockener Straße, das andere enthielt Fahrgeräusche von Fahrzeugen im Regen. Alarmsirenen und Verkehrsgeräusche wurden in disjunkte Trainingsdaten (250 Samples) und Testdaten (128 Samples) unterteilt, so dass der entwickelte Erkenner nur mit Signalen getestet wurde, die nicht bereits im Training verwendet wurden. In den Trainingsdaten wurde zunächst ausschließlich das normale Verkehrsgeräusch als Störgeräusch verwendet. Alarmsirenen und Verkehrslärm wurden in verschiedenen Signal-Rausch-Abständen (Engl. signal-to-noise ratio, SNR) addiert. Das Training des Erkenners erfolgte als sogenanntes Multi-Kondition-Training bei SNR von -5 bis +10 dB in Schritten von 5 dB. Für jedes der 250 Trainingssignale wurde zufällig ein SNR ausgewählt. Alle Signale wurden mit einer Samplingrate von f_s = 16 kHz abgetastet.

2.2 Ereigniserkenner

Zur automatischen Detektion und Klassifikation der Alarmsignale im Verkehrslärm wurde ein Erkenner basierend auf Hidden Marko Modellen (HMMs, [8]) verwendet, d.h. ein statistisches Verfahren, welches auf Beobachtungen beruht, die von nicht beobachtbaren („hidden“) Zuständen erzeugt werden. In akustischen Anwendungen entsprechen diese Zustände typischerweise Zeitintervallen eines bestimmten Signaltyps (z.B. Silben oder Phoneme bei der automatischen Spracherkennung), während die Beobachtungen aus bestimmten Merkmalen bestehen, die aus den Audiosignalen berechnet werden können. In der Trainingsphase wurden aus den annotierten Trainingsdaten die am besten geeigneten Merkmale zur Unterscheidung zwischen Ziel- und Hintergrundsignalen ausgewählt. In der Testphase wurden die Merkmale für die unbekannten Signale berechnet und eine Klassifikation in „Alarmsignal vorhanden“ bzw. „kein Alarmsignal vorhanden“ anhand der Ähnlichkeit der trainierten Merkmale vorgenommen.

Drei verschiedene Arten von Merkmalen wurden getestet: *(i)* Mel Frequency Cepstral Coefficients (MFCC, [8]), *(ii)* Merkmale basierend auf der Autokorrelationsfunktion (im Folgenden abgekürzt mit auto correlation function, ACF) [4, 5] sowie *(iii)* eine Mischung aus MFCC und Autokorrelationsmerkmalen (im Folgenden abgekürzt mit MFiFFT). Alle Merkmale wurden in kurzen Zeitfenstern (Länge 25 ms, Überlapp 10 ms) berechnet.

3 Ergebnisse

3.1 Einfluss der Merkmale

Der Vergleich der drei unterschiedlichen Merkmale ist in **Bild 1** dargestellt, welches die Genauigkeit (Engl. accuracy) der Erkennung in Abhängigkeit des Eingangs-SNR zeigt. Die Genauigkeit ist dabei der Anteil der korrekt als Alarmsirene bzw. Verkehrsgeräusch klassifizierten Signale. Unterschiedliche Symbole repräsentieren die verschiedenen Merkmale, während die beiden Hintergrundgeräusche durch durchgezogene (normales Verkehrsgeräusch) bzw. gestrichelte (Fahrgeräusche bei Regen) gekennzeichnet sind.

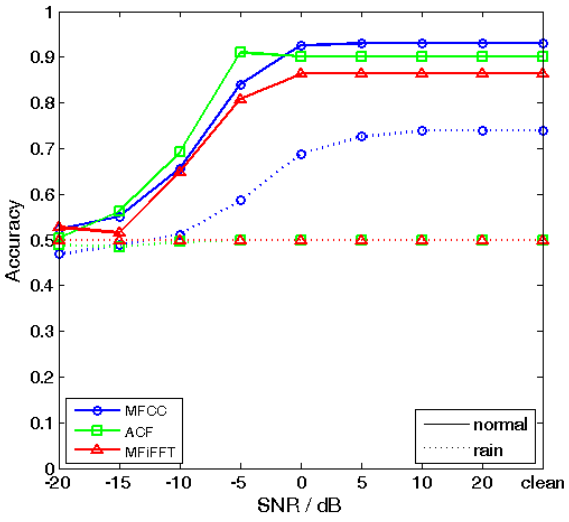

Bild 1. Genauigkeit der Ereigniserkennung über dem SNR für unterschiedliche Merkmale und Störgeräusche.

Es zeigt sich, dass für alle Merkmale und normales Verkehrsgeräusch der erwartet ansteigende Trend bei besser werdenden SNR auftritt. Bei sehr niedrigem SNR entspricht die Genauigkeit in etwa der Ratewahrscheinlichkeit von 0,5. Bei einem SNR ≥ -5 dB erreicht der Erkenner Genauigkeiten von > 0,8, welche jedoch auch bei sehr hohen SNR vor allem aufgrund von Fehlalarmen nicht den Maximalwert erreichen. Beste Erkennungsleistungen werden für SNR > 0 dB mit MFCC Merkmalen erreicht. Für alle Merkmale sind die Genauigkeiten für normales Verkehrsgeräusch besser als für die Hintergrundgeräusche

im Regen. Dies ist dadurch begründet, dass Letztere im durchgeführten Experiment nicht im Trainingsmaterial enthalten und somit für die Erkenner in der Testphase nicht bekannt waren. Gegenüber dem nicht trainierten Hintergrundgeräusch zeigen sich wiederum die MFCC Merkmale als deutlich robuster, d.h. der Einbruch der Erkennungsleistung ist geringer als für die anderen Merkmale, mit welchen für das nicht trainierte Störgeräusch keine Klassifikation mehr möglich ist (alle Werte nahe der ratewahrscheinlichkeit).

3.2 Einfluss des Spektrums

Die in Abschnitt 3.1 gezeigten Ergebnisse wurden ohne weitere Annahmen über die physikalischen Eigenschaften der Ziel- und Hintergrundgeräusche ermittelt. Alarmsignale zeichnen sich jedoch durch ihren tonalen Charakter aus, der im Spektrum zu deutlich ausgeprägten Obertönen und somit zu hochfrequenten Signalanteilen führt, während die Energie der Verkehrsgeräusche überwiegend bei tieferen Frequenzen liegt. Daher ist eine Steigerung der Erkennungsleistung zu erwarten, wenn der der Klassifikation zugrunde liegende Frequenzbereich eingeschränkt wird. Um diese spektrale Abhängigkeit zu untersuchen, wurden basierend auf den MFCC Merkmalen unterschiedliche Bandpassbegrenzungen vorgenommen. Im folgenden Experiment wurden Bandbegrenzungen von 620 bis 8000 Hz sowie von 1000 bis 4500 Hz untersucht. Die erste Bandbegrenzung entspricht einem Ausschluss des tiefen Frequenzbereiches, in dem das Verkehrsgeräusch den Großteil seiner Energie enthält. Die zweite Bandbegrenzung betont die am deutlichsten ausgeprägten Obertöne der Alarmsignale. Zum Vergleich sind die Ergebnisse bei voller Bandbreite (0-8000 Hz) ebenfalls nochmals in **Bild 2** dargestellt.

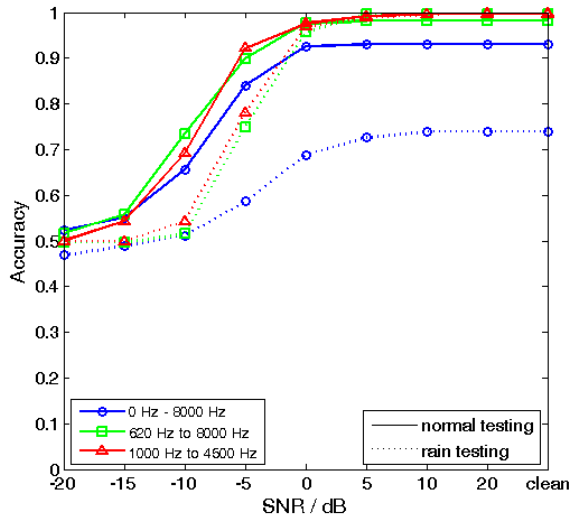

Bild 2. Genauigkeit der Erkennungsleistungen für unterschiedliche Bandpassbegrenzungen und Störgeräusche.

Die Begrenzung des Frequenzbereiches führt zu einer deutlichen Erhöhung der Erkennungsleistungen und führt bei hohen SNR zu nahezu fehlerfreier Klassifikation für normale Verkehrsgeräusche. Selbst bei einem SNR von -5 dB übersteigt die Erkennungsleistung eine Genauigkeit von 0,9. Die größte Verbesserung der Erkennungsleistung erfolgt für das nicht im Training enthaltene Regengeräusch, für das ebenfalls Genauigkeiten > 0,9 bei SNR von 0 dB oder höher erreicht werden. Dies zeigt, dass die Robustheit der Erkennung gegenüber unbekannten Hintergrundgeräuschen durch eine passende Auswahl des Frequenzbereiches deutlich gesteigert werden kann.

4 Diskussion

In dieser Studie wurde ein auf HMMs basiertes System zur automatischen Unterscheidung von Alarmsirenen und Verkehrsgeräuschen implementiert und mit realistischen Signalen validiert. Die dabei erzielten Ergebnisse zeigen, dass bis zu einem SNR von 0 dB eine nahezu optimale Klassifikationsgenauigkeit erreicht werden kann, wenn die entsprechenden Merkmale sowie der betrachtete Frequenzbereich geeignet gewählt werden. Die erzielten Erkennungsleistungen liegen damit oberhalb bisher in der Literatur beschriebener Ergebnisse (z.B. [5]) und verdeutlichen, dass eine automatische akustische Ereigniserkennung prinzipiell für verkehrsrelevante Szenarien anwendbar ist. Eine weitere Verbesserung der Erkennergenauigkeit oder Robustheit des Systems könnte bspw. durch gezielteres Training, eine weitere Optimierung des untersuchten Frequenzbereiches, die Verwendung weiterer Merkmalstypen oder den Einsatz von Audiosignalverbesserungsalgorithmen (z.B. zur Störgeräuschreduktion, siehe z.B. [10]) erreicht werden, was jedoch in dieser prototypischen Studie nicht untersucht wurde.
Das hier vorgestellte Szenario zur Erkennung von Alarmsirenen ist nur ein Beispiel möglicher Anwendungen von akustischen Erkennertechnologien im Automobil. Weitere Möglichkeiten wären das Detektieren und Klassifizieren von plötzlich auftretenden Fahrzeuggeräuschen (Klappern, Quietschen, etc.) oder das Erkennen des Fahrbahnbelages anhand der Rollgeräusche. Durch den Einsatz mehrerer Mikrofone könnte dabei neben der Erkennung und Klassifikation der akustischen Ereignisse auch eine Lokalisation der Schallquellen erfolgen. Insgesamt deutet diese Studie darauf hin, dass die akustische Modalität eine vielversprechende Ergänzung zu bestehenden Sensoren für die Gefahrenerkennung und Diagnose im Automobil ist.

5 Literatur

[1] C. Müller, A. Grünewald, T. Bayer, M. Habbaba, S. Rose, S. Schalk, D. Zimmermann, M. Mielke, A. Schäfer und R. Brück (2010), “ASD - automatic siren detection - a mixed signal ASIC for detection of emergency signals in general traffic situations,” in University Booth DATE 2010, Dresden, Germany, 2010.

[2] N. Vidyasagar, A. Jain und M. Bianchi (2010), “Emergency vehicle detection system,” Universität

Illinois, Tech. Rep., Dezember 2010, eCE 445 Senior Design Project Fall 2010, Project 15.

[3] X. Xiao und H. Yao (2009), “Automatic detection of alarm sounds in cockpit voice recordings”, IITA International Conference on Control, Automation and Systems Engineering (CASE), S. 599-602, Zhangjiajie, China, 2009.

[4] M. Mielke, A. Schäfer, M. Wahl, and R. Brück (2010), “Integrated circuit for detection of acoustic emergency signals in road traffic,” in International Conference Mixed Design of Integrated Circuits and Systems, MIXDES 2010, Breslau, Polen, 2010.

[5] R. A. Lutfi and I. Heo (2012), “Automated detection of alarm sounds,” Journal of the Acoustical Society of America, vol. 132, no. 2, EL 125-EL128, September 2012.

[6] F. Beritelli, S. Casale, A. Russo und S. Serrano (2006), “An automatic emergency signal recognition system for the hearing impaired,” in 12th Digital Signal Processing Workshop, Wyoming, USA, September 2006.

[7] D. Ellis (2001), “Detecting alarm sounds,” in Proceedings of Recognition of real-world sounds: Workshop on consistent and reliable acoustic cues, Aalborg, Dänemark, S. 59-62.

[8] L. R. Rabiner (1989), “A tutorial on hidden Markov models and selected applications in speech recognition,” in Proceedings of the IEEE, 1989, S. 257–286.

[9] L. Rabiner L. und B.-H. Juang (1993). „Fundamentals of speech recognition," Prentice-Hall Signal Processing Series, Englewood Cliffs, USA.

[10] P. Vary, U. Heute und W. Hess (1988), „Digitale Sprachsignalverarbeitung“, Teubner, Stuttgart, Deutschland.