

Sound- und Mustererkennung unter Nutzung von Spracherkennungs-Technologien

Diane Hirschfeld *, Rüdiger Hoffmann **, Peter Holstein, Michael Koch***

*voice INTER connect GmbH, Dresden, **Institut für Akustik und Sprachkommunikation der TU Dresden,

***Sinus Messtechnik GmbH, Leipzig

Email: hirschfeld@voiceinterconnect.de

Für vielfältige Klassifikations- und Diagnoseaufgaben (Qualitätskontrolle von Bremsbelägen, Überwachungsaufgaben z.B. Detektion von Störzuständen in Getrieben, Sound-Design technischer Geräte und Kfz, usw.) werden heutzutage akustische Signale eingesetzt. Die eigentliche Aufgabe der Klassifikation ist jedoch zumeist noch an menschliche Experten gebunden, die mit langjähriger Erfahrung die zu beobachtenden Signale auditiv bewerten und daraus auf den Systemzustand schließen können. Technische Systeme sind für solche Klassifikationsaufgaben sehr gut geeignet, da sie hinsichtlich der Merkmalsdetektion beliebig genau an das Problem angepaßt werden können, reproduzierbare Ergebnisse liefern und keine Ermüdungserscheinungen aufweisen. Unter bestimmten Bedingungen ist eine objektivierte Klassifizierung und Erkennung auch unter akustisch rauen Bedingungen von großer praktischer Relevanz.

Einführung

Im technischen oder biologischen Bereiche gibt es viele Klassifikationsaufgaben. Oft ist es ausreichend, die Aktivität einer bestimmten Schallquelle zu detektieren. Dies kann im Vergleich zur Spracherkennung als eine recht einfach zu realisierende Fragestellung erscheinen, die sich aber sofort erheblich verkompliziert, wenn man Signale in Anwesenheit beliebiger Fremdgeräusche zu finden versucht.

Der vorliegende Beitrag beschreibt erste Aktivitäten zur echtzeitnahen Sofortbewertung der im akustischen Signal verborgenen Informationen. Bei Überwachungsvorgängen werden oft große Datenmengen aufgezeichnet, deren Auswertung aufgrund hohen manuellen Aufwandes nicht praktikabel ist. Tatsächlich sind für die Überwachung nur jene Daten von (Aufzeichnungs-) Interesse, die wirklich die gesuchten Ereignisse enthalten, und dort auch nur bestimmte (gegenüber den mit voller Datenbreite aufgezeichneten Audiosignalen deutlich reduzierten) informationstragende Merkmale.

Klassifikation von Bremsgeräuschen

In der Industrie erfolgt die Bewertung von Bremseigenschaften entweder im Fahrversuch durch objektive Messungen und Merkmalsextraktion bzw. subjektive Beurteilung der Bremsgeräusche [1] oder durch Messungen am Bremsenprüfstand, die im folgenden die Hauptrolle spielen sollen.

Das Verhältnis subjektiver zu objektiver Bewertung beträgt üblicherweise 80% zu 20%. Ein ausführliches Training ist bisher unabdingbar für eine subjektive Bremsgeräuschbewertung.

Klassifikationsziel

Es soll eine Datenerfassungs- und Bewertungsstrategie für Bremsgeräusche sowie eine intelligente problemangepaßte Datenreduktion und Informationsextraktion als Beitrag zur Prüfstandsautomatisierung entwickelt werden. Das zu entwickelnde Verfahren soll zunächst eingesetzt werden als Ereignisdetektor, der

auf ein Schallereignis hin die Meßwerterfassung einschaltet. Später soll das Verfahren auch diagnostische Rückschlüsse vom Geräusch auf die vermutete Geräuschquelle und -ursache im Fahrversuch ermöglichen, auch wenn viele Nebengeräusche (Wind, Roll- und Fahrwerksgeräusche etc.) vorhanden sind.

Datenerfassung am Prüfstand

Die Erfassung der Audiosignale erfolgte durch Aufnahme der Bremsgeräusche am Bremsenprüfstand mittels Freifeldmikrophon radial zur Bremsscheibe in 100 mm Abstand mit Harmony [3]. Für die hier beschriebenen Experimente wurden 26 Schallereignisse aufgenommen, und in wav-Dateien bei 51.2 kHz, unkodiert in 16 Bit und zweikanalig abgespeichert. Die nachfolgenden Untersuchungen beziehen sich ausschließlich auf diese aufgenommenen Schallereignisse.

Subjektive Klassifikation der Schallereignisse

Zunächst wurden alle Signale in hochfrequente und niederfrequente Geräusche unterteilt und weitere Beschreibungsattribute ausgewählt. Eine Überprüfung und Zuordnung der Attribute auf die Signale erfolgte im Hörversuch. Dazu wurden die am Prüfstand gewonnenen Geräusche in zufälliger Reihenfolge über Kopfhörer ohne Kenntnis der Randbedingungen (Drehzahl, Druck usw.) abgespielt. Es wurden ausschließlich vorgegebene Attribute zugeordnet. Anschließend wurde durch Korrelation subjektiver Bewertungen und objektiver Messungen eine Unterteilung der Geräusche in Ereignisklassen durchgeführt (Tabelle 1).

Tabelle 1 - Subjektive Klasseneinteilung von Bremsgeräuschen

Klasse	Besonderheit	Bremsdruck	Klang
1	Zerstörter Belag	8 bar	Hell, schleifend, tonal
2	Zerstörter Belag	50 bar	Schleifend, breitbandig
3	niederfrequent	5 bar	Schwebend, schrappend
4	hochfrequent	5 bar	schwingend, quietschend

Automatische Erkennung

Ganz allgemein ist ein beliebiger Mustererkenner aus den beiden Komponenten Analysator bzw. Klassifikator aufgebaut. Die Analyse dient der Transformation des Eingangssignals in eine für die Klassifikation geeignete Repräsentation (Merkmalsberechnung) bei gleichzeitiger Datenreduktion. Der Klassifikator trifft eine Klassifikationsentscheidung auf der Basis seines Referenzwissens über die zu unterscheidenden Klassen. Für die Klassifikation zeitlich veränderlicher Muster, sowie die Ereignisdetektion und die Unterdrückung von Störsignalen stellt die Sprachverarbeitung verschiedene erprobte Verfahren und Algorithmen zur Verfügung.

Für das Training des Erkenners müssen die zu klassifizierenden Daten in Trainings- und Teststichprobe unterteilt werden. Mit der Trainingsstichprobe wird das Referenzwissen des Erkenners erzeugt. Mit der Teststichprobe wird die Klassifikationsgenauigkeit

überprüft. Durch veränderte Parametrierung lassen sich die Algorithmen des Spracherkenners auf beliebige Klassifikationsaufgaben anpassen.

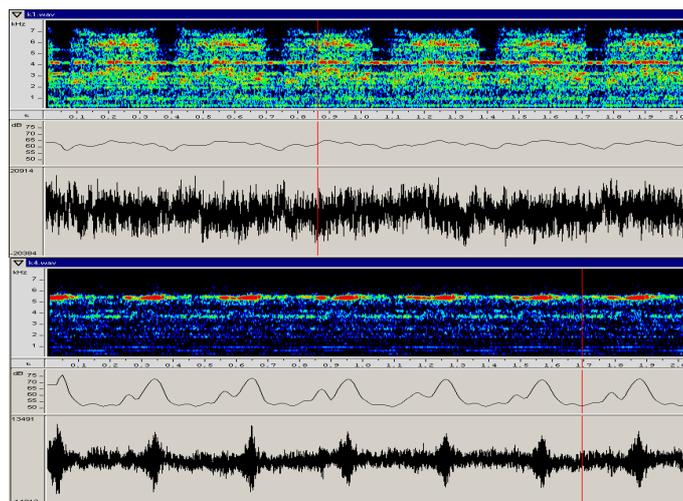


Abbildung 1 - Spektrum, Energieverlauf und Zeitsignal von Bremsgeräuschen. Oben: Klasse 1, Unten: Klasse 4.

Algorithmus

Eingehende Untersuchungen der spektralen Zusammensetzung und der Merkmalsvergleich zwischen den 4 identifizierten Klassen erbrachten, daß die zu untersuchende Bandbreite der Bremsignale auf 8 kHz begrenzt werden kann. Daher wurden alle Testsignale auf 16 kHz Abtastfrequenz, sowie auf einen Kanal reduziert. Die Extraktion der Merkmale erfolgt durch Analyse von Signalzeitfenstern von 32 ms Länge mit einer Folgefrequenz von 100 Hz.

Zeitfunktionsbasierte Merkmale. Ein Versuch der Korrelation auditiver und meßbarer Ereignisse erbrachte, daß Schrapfen und Schwebungen von Amplitudenmodulationen herrühren, die sich deutlich im zeitlichen Verlauf der Signalenergie widerspiegeln. Eine Auswertung dieser Merkmale wurde jedoch nicht weiter verfolgt.

Frequenzbasierte Merkmale. Schwingungen und Rotationsvorgänge sind im Frequenzbereich durch einzelne deutliche Spektralkomponenten charakterisiert. Bei Abnutzung von Bremsbelägen erzeugt z.B. die Reibung von Stahl auf Stahl typische Schwingungen, die durch Detektoren (z.B. Bandpaß-Filter mit nachgeschaltetem Schwellwertschalter) aufgefunden werden können. Sofern es sich nicht um Auslaufvorgänge handelt, können die untersuchten Bremsgeräusche als stationär angesehen werden. Aufgrund der Verteilung der beobachteten Merkmale (Abbildung 1) wurde nach Durchführung der Fouriertransformation das Spektrum im Frequenzbereich von [100..8000] Hz in 24 linearen Filterkanälen zusammengefaßt. Die Grenzfrequenz des untersten Frequenzkanals wurde auf 100 Hz festgelegt. Daraus ergibt sich eine Kanalbreite jedes Filterkanals von 330 Hz. Nach Normierung des Filterbankspektrums auf die mittlere Energie und zeitlicher Mittelung der normierten Filterbankspektren wurden die beiden größten Maxima bestimmt und die Indizes der dazugehörigen Frequenzkanäle ermittelt. Diese Indizes wurden für die Klassenentscheidung herangezogen (Abbildung 2).

Ergebnisse

Für komplexere Signale mit zeitlicher Struktur, wie zum Beispiel für Sprachsignale, Schnarchen bzw. Vogelgesang (Abbildung 3)

kommen komplexe Klassifikatoren wie z.B. DTW oder HMMs zum Einsatz, die einen Mustervergleich bei gleichzeitiger optimaler Zeitanpassung der beiden zu vergleichenden Muster durchführen. Für die Detektion stationärer Ereignisse eignen sich insbesondere Algorithmen, die in der Sprachverarbeitung zur Ermittlung von Äußerungsgrenzen bzw. zur Sprach-Pause-Detektion (Voice activation detector – VAD) verwendet werden. Diese werten Energieänderungen im Filterbankspektrum der Primäranalyse aus, und wurden für das folgende Experiment leicht angepaßt.

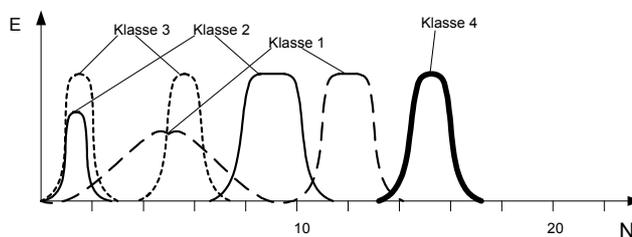


Abbildung 2 – Merkmalsverteilung für erstes und zweites Energiemaximum im Histogramm

Zur Klassifikation der Filterenergien wurden die in Abbildung 3 dargestellten Fuzzy-Bewertungen verwendet. Die damit erreichbare Erkennungsrate¹ von 96.1 % zeigt, daß diese einfache Vorgehensweise eine robuste Klassifikation erlaubt.

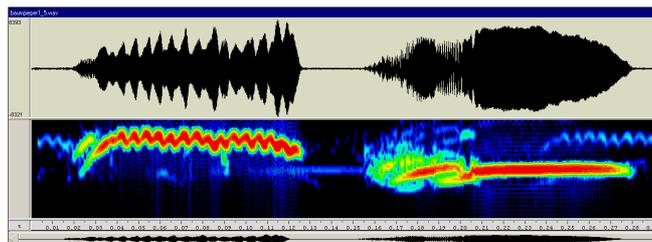


Abbildung 3 - Zeitfunktion und Spektrogramm Vogelgesang. Baumpieper.

Ausblick und Danksagung

Weitere Experimente zielen auf die Klassifikation von Biosignalen (z.B. Vogelstimmen, Schnarchen) für die Entwicklung automatischer Überwachungstechnik.

Ein herzlicher Dank gebührt Herrn Dipl.-Ing. Bader, Technische Universität Ilmenau, für die bereitgestellten Daten und Informationen zum Prüfverfahren und zum Prüfstand.

Literatur

- [1] Rumold, W., Keiper, W. „Bremsgeräusche – Vergleich subjektiver und objektiver Bewertungen“, in: Syed R. Ahmed: Akustik und Aerodynamik des Kraftfahrzeuges, Expert-Verlag, 1995.
- [2] Holstein, P., Koch, M., Hirschfeld, D., Hoffmann, R., Bader, D., Augsburg, K.: „A strategy for signal recognition under rough conditions“, Proceedings INTER-NOISE 2003, Korea, 2003
- [3] P. Holstein, R. Müller, A. Uhlmann: „Ein modulares Konzept für Mess-Systeme unter MATLAB“, Proceedings 5. Tagung Technische Diagnose, Fachhochschule Merseburg, 2002.

¹ Erkennungsrate: korrekt erkannte Realisierungen, bezogen auf die Gesamtzahl an Realisierungen