

Einkanalige Detektion von Windgeräuschen mittels Klassifikation

Ivo Beilfuss¹, Jacob Aderhold², Jörg Bitzer¹

¹ Institut für Hörtechnik und Audiologie, Jade Hochschule Oldenburg, E-Mail: ivo.beilfuss@jade-hs.de

² Universität Oldenburg, E-Mail: jacob.aderhold@uni-oldenburg.de

Einleitung

Windgeräuschen können sowohl zur Verschlechterung der Klangqualität als auch zu einer Verringerung des Sprachverstehens führen. Algorithmen zur Reduktion von Windgeräuschen existieren dabei bereits in Echtzeit verarbeitenden Systemen (z.B. Hörgeräte, Handys) und in Offline-Systemen (z.B. Audio-Restauration, Nachbearbeitung im Tonstudio). Diese können jedoch bei ungestörtem Audiomaterial Artefakte erzeugen und profitieren daher von einer vorgeschalteten Detektionsstufe. Aufgrund ihrer guten Detektionsleistung wurden in der Vergangenheit zwei- und mehrkanalige Detektions- und Reduktionsalgorithmen entwickelt [2]. In einigen Geräten und Situationen ist jedoch nur ein Aufnahmekanal vorhanden (Camcorder, Reportagen, Hörgeräte). Da einkanalige Ansätze [1] in der Literatur unterrepräsentiert sind, ist das Ziel dieses Beitrags die Entwicklung einer einkanaligen Detektionsstufe für Windgeräusche.

Eigenschaften von Windgeräuschen

Bei Außenaufnahmen kann Wind ein tieffrequentes Störgeräusch erzeugen. Dies geschieht durch turbulente Luftströmungen am Mikrofon [3]. Weiterhin können durch Übersteuerung des Vorverstärkers Verzerrungen entstehen. Windgeräusche besitzen bis ca. 300 Hz ein relativ flaches Langzeitspektrum gefolgt von einem Abfall mit 9 dB/Oktave [7].

Implementierung

Das Problem der Windgeräuschdetektion soll mit Hilfe eines Standardklassifikationssystems mit den zwei Klassen *Windgeräusch* und *Sprache* gelöst werden: Aus einer vorhandenen Trainingsdatenbank ist eine Liste an Merkmalen zu extrahieren, um damit anschließend einen Klassifikator zu trainieren. So kann dieser für Testdaten, deren Klassenzugehörigkeit unbekannt ist, eine blockweise Entscheidung über die Klassenzugehörigkeit treffen. Wesentliche Aspekte der Optimierung eines solchen Systems ist somit das Finden und die Auswahl der zu verwendenden Merkmale und des geeignetsten Klassifikators.

Datenbank

Die Trainingsdatenbank bestand insgesamt aus 64 Minuten Windgeräuschaufnahmen mit teilweise geringen Sprachanteil und 62 Minuten Sprachaufnahmen (arabisch, deutsch, hebräisch, japanisch, mandarin, spanisch, vietnamesisch). Das Audiomaterial besaß verschiedene Aufnahmequalitäten (Handyaufnahmen, Studioaufnahmen, etc.) und lag einkanalig, mit einer Samplingrate von 44,1 kHz und 16 Bit Quantisierung im WAV-Format vor.

Merkmale

Die in diesem Beitrag verwendeten Merkmale setzen sich aus Grund-Merkmalen, die direkt aus Audio-Daten berechnet werden und Zeitreihenanalysen zusammen. Die Zeitreihenanalyse verwendet gleitende Analysefenster, wodurch neue Merkmale entstehen, welche die gleiche Zeitbasis besitzen.

Varianz der MFCCs:

Das erste Merkmal basiert auf den *Mel Frequency Cepstral Coefficients* (MFCCs). Die Implementierung erfolgt nach [4]. Zwölf MFCCs werden hier als Grund-Merkmal zwischen 50 und 8000 Hz und einer Blocklänge von 50 ms extrahiert. Das verwendete Merkmal berechnet sich aus der Varianz jeder MFCC-Dimension über 2000 ms.

Mittelwert und Varianz des spektralen Zentroids:

Zunächst ist der *spektrale Zentroid* [6] über 50 ms zu extrahieren. Aus diesem können die zwei oben genannten Merkmale durch Zeitreihenanalyse des Mittelwerts und der Varianz jeweils über 1000 ms generiert werden.

Mittelwert und 95%-Perzentil der Schärfe:

Zwei weitere Merkmalsdimensionen setzen sich aus dem Grund-Merkmal *Schärfe* [6] zusammen, das in 50 ms Blöcken zu extrahieren ist, und aus der anschließenden Berechnung des Mittelwerts zum einen, sowie der Berechnung des 95%-Perzentils zum anderen mit einer Analyselänge von je 1000 ms.

Verschiedene Zeitreihenanalysen der Varianz logarithmierter Frequenz-Bins:

Dieses Merkmal wurde durch Messergebnisse in [7] motiviert, in denen Langzeitspekten von Windgeräuschen einen relativ flachen Verlauf bis ca. 300 Hz besitzen. Die Varianz der zuvor logarithmierten Bins des Betragsspektrums im Bereich von 100 Hz bis 300 Hz stellen das Grund-Merkmal und somit ein Maß für die Flachheit dar, das über 100 ms zu extrahieren ist. Aus diesem Grund-Merkmal lassen sich nun zwei Merkmale ableiten, die in dieser Arbeit Verwendung finden und die beide zwei aufeinanderfolgende Zeitreihenanalysen erfordern. Für das erste Merkmal ist die Standardabweichung über eine Analyselänge von 1000 ms zu berechnen, gefolgt vom Mittelwert über 2000 ms. Das Zweite setzt sich aus dem 95%-Perzentil des Grund-Merkmals zusammen, dass über 1000 ms berechnet wird, ebenfalls gefolgt von dem Mittelwert über 2000 ms.

Evaluation

Zunächst erfolgte eine Selektion der im vorrangegangenen Abschnitt erwähnten Merkmalsdimensionen durch

das mRMR-Kriterium [5]. Durch diese Statistik-basierte Selektionsmethode kann für verschiedene Dimensionszahlen eine nach dem Kriterium optimale Kombination von Merkmalsdimensionen erfolgen. Die Selektion des Klassifikators erfolgte über einen Round-Robin-Test mit der Trainingsdatenbank. Diese wurde dafür zufällig permutiert und gedrittelt, so dass in drei Durchgängen ein Drittel als Trainingsmaterial und zwei Drittel als Pseudo-Testmaterial verwendet werden konnte. Die Selektion des Klassifikators erfolgte durch den Vergleich der mittleren Genauigkeit (richtig erkannte Blöcke durch Anzahl vorhandener Blöcke). Folgende Klassifikationsmethoden wurden getestet:

- Gaussian Mixture Models (GMM) mit Diagonalmatrix, Anzahl der Modelle = {2, 4, 8, 16, 32}
- Vektor-Quantisierer mit Mahalanobis-Distanz, Anzahl Vektoren = {8, 16, 32, 64}
- k-Nearest-Neighbor (KNN) mit Mahalanobis-Distanz, $k=\{3, 7, 11, 15, 19\}$
- Maximum a-posteriori (MAP) / naive-Bayes

In Abbildung 1 sind die Ergebnisse der selektierten Merkmalskombinationen mit verschiedenen Dimensionszahlen dargestellt. Zum Vergleich wurde für jede Dimensionszahl das Ergebnis der Klassifikationsmethode mit der höchsten gemittelte Genauigkeit ausgewählt.

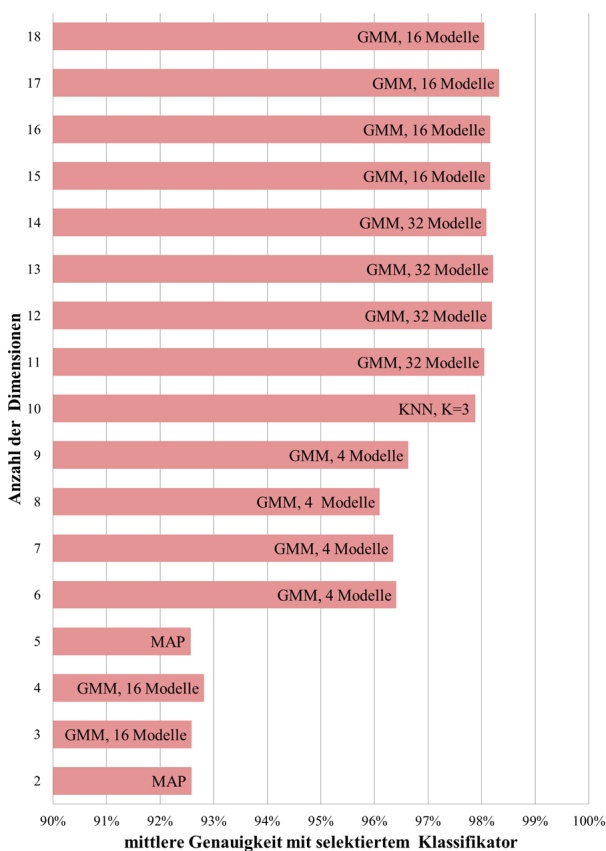


Abbildung 1: Mittlere Genauigkeit des selektierten Klassifikators für verschiedene Merkmalskombination. Für jede Dimensionszahl wurde zuvor eine Merkmalskombination selektiert.

Es zeigt sich insgesamt eine hohe Klassifikationsleistung, selbst bei einer geringeren Anzahl von Merkmalsdimensionen und ein Anstieg der mittleren Genauigkeit zu großen Dimensionszahlen hin. Ab zwölf Merkmalsdimensionen ist eine Sättigung in den mittleren Genauigkeiten zu beobachten. Bei den Klassifikationsmethoden dominieren die GMMs bis auf wenige Ausnahmen, jedoch steigt die Anzahl der Modelle nicht zwangsläufig mit der Dimensionszahl und ist somit nicht intuitiv vorhersagbar.

Zusammenfassung und Fazit

Die Ergebnisse des Round-Robin-Test lassen darauf schließen, dass eine einkanale Detektion von Windgeräuschen über einen Klassifikationsansatz möglich ist. Die Hinzunahme von Zeitreihenanalysen mit gleitenden Analysefenstern, um Grund-Merkmale hinsichtlich bestimmter Eigenschaften zu untersuchen, bewirkt eine stärkere Klassentrennung im Merkmalsraum. Dies setzt jedoch voraus, dass eine mehrere Sekunden dauernde Windgeräuschstörung vorliegt. In Zukunft sollten weitere Merkmale entwickelt werden, die einerseits einen größeren Raum von Eigenschaften der Klasse Windgeräusche abdecken und andererseits mit kürzeren Analysefenstern auskommen. Das Einbinden weiterer Klassifikationsmethoden in das System ist ebenso angedacht.

Literatur

- [1] Kates, J. M., A hearing aid with suppression of wind noise, U.S. Patent 20 070 030 989, Februar, 2007.
- [2] Franz, S., Bitzer, J., 2010, Multi-channel algorithms for wind noise reduction and signal compensation in binaural hearing aids, IWAENC Proceedings.
- [3] Dillon, H., 2001, Digital hearing aids, Thieme.
- [4] Niemann, H., 1983, Klassifikation von Mustern, Vol. 1, Springer, Heidelberg.
- [5] Peng, H., Long, F., Ding, C., 2005, Feature Selection Based on Mutual Information: Criteria of Max-Dependency, Max-Relevance, and Min-Redundancy. In: IEEE Transactions on pattern analysis and machine intelligence, vol. 27, no. 8 (2005)
- [6] Peeters, G., 2004, A large set of audio features for sound description (similarity and classification) in the CUIDADO project, Report, Analysis/synthesis team, Paris Institut de Recherche et Coordination Acoustique/Musique (IRCAM)
- [7] Larsson, P., Olsson, P., 2004, Detection of wind noise in hearing aids, Masters thesis, Department of Electrosience, Lund Institute of Technology

Diese Arbeit wurde gefördert durch das BMBF (Förderkennzeichen 03FH030PX2). Den Inhalt und die Erkenntnisse verantworten ausschließlich die Autoren.