# Blinde Schätzung des globalen Signal-Rauschabstands und perzeptiv motivierte Alternativen bei Sprache und Musik

Jan Willhaus, Joerg Bitzer, Jens-Alrik Adrian
*Institut für Hörtechnik+Audiologie, 26121 Oldenburg, Deutschland, Email: jan.willhaus@student.jade-hs.de*

## Einleitung

Das Signal-Rauschverhältnis (SNR), im Weiteren als logarithmisches Maß als Signal-Rauschabstand verwendet, ist ein wichtiges Beschreibungsmerkmal akustischer Szenen und von zentraler Bedeutung als Signaleigenschaft. Häufig wird der globale SNR verwendet, der durch den Bezug des Rauschenanteils am Gesamtsignal beispielsweise eine begrenzte Aussage über die Aufnahmequalität zulässt. Entrauschungsalgorithmen (*Denoising*) können den globalen SNR deshalb als Entscheidungshilfe und Steuergröße verwenden, da die Algorithmen weder bei sehr gutem SNR noch bei sehr geringem SNR eine Verbesserung der Verständlichkeit und Qualität durch die Entrauschung erzielen bzw. durch Artefakte die Signalqualität verschlechtern. Insbesondere die maximal anzuwendene Geräuschreduktion, als Parameter oft *Spectral Floor* bezeichnet, kann durch eine gute SNR Schätzung besser an das vorliegende Signal angepasst werden. Die Schätzung des globalen SNR ist jedoch in der Regel ungenau, vor allem bei längeren, instationären oder spektral gefärbten Signalen. Die hier vorgestellte Analyse des Signal-Rauschabstands in Frequenzbändern und Zeitblöcken soll diese Ungenauigkeit vermindern und ermöglicht es, eine Färbung des Rauschens sowie des Signals bei der Denoising-Entscheidung besser berücksichtigen zu können.

## Algorithmus

Die Grundlage des hier vorgestellten Verfahrens ist die spektrale Zerlegung des Signals mittels Gammatone-Filterbank mit anschließender Perzentilanalyse des 2. und 98. Perzentils des Spektrogramms in blockweiser Verarbeitung über etwa 5 Sekunden (siehe Abbildung 1). Nach Anwendung einer zuvor aus A-priori-Wissen ermittelten Wertekorrektur (*Mapping*, siehe nächster Abschnitt) ergibt sich so ein frequenzabhängiges SNR-Profil pro Block, das den Anteil der Rauschleistung im einzelnen Band widerspiegelt und somit auch Rückschlüsse auf die Färbung – und bei Vergleich der Blöcke untereinander auch auf die Stationarität – des Rauschens zulässt. Grundlage des A-priori-Wissens ist der frequenzabhängige globale SNR, kurz FSNR[1].

## Mapping des Perzentilverhältnisses

Bei der Analyse von Sprachmaterial und der Verwendung des Perzentilverhältnisses des 2. und 98. Perzentils stellt sich für negative SNR eine Sättigung der Relation ein. Zur Untersuchung wurde zunächst sprachpausenmoduliertes weißes Rauschen und unmoduliertem Rauschen bei bekanntem SNR gemischt und der Analyse zugeführt. Aus den Werten kann eine Kurve der notwendigen Wertekorrektur (*Mapping-Curve*) generiert werden. Deren Verlauf zeigt zum Teil Abhängigkeiten vom prozentualen Anteil der Sprachpausen am Signal (siehe Abbildung 2). Zur Bestimmung einer allgemeingültigen Mapping-Curve wurde deshalb eine breite Auswahl von Sprachmaterial verwendet. Dieses setzt sich aus gelesener und spontaner Sprache zusammen, die von 22 Sprechern bei $f_s = 48\,\mathrm{kHz}$ (24 bit) aufgezeichnet wurde. Insgesamt liegen etwa 120 Minuten Sprache mit einem Pausenanteil von 15,4 % zugrunde.

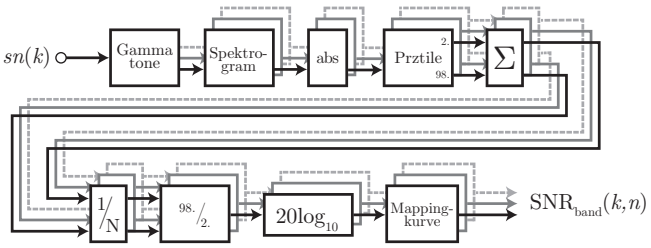



**Abbildung 1:** Blockschaltdiagramm des vorgeschlagenen Algorithmus zur Schätzung des SNR auf Basis der bandweisen Perzentilanalyse. Nach der Gammatone-Filterung erfolgen alle weiteren Schritte in Bändern; das Ergebnis ist pro Eingangsblock $sn(k)$ ein SNR-Profil $\mathrm{SNR}_{\mathrm{band}}(k,n)$.

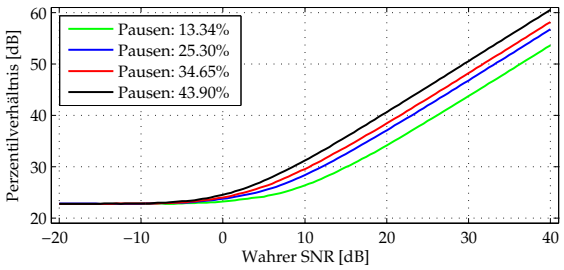
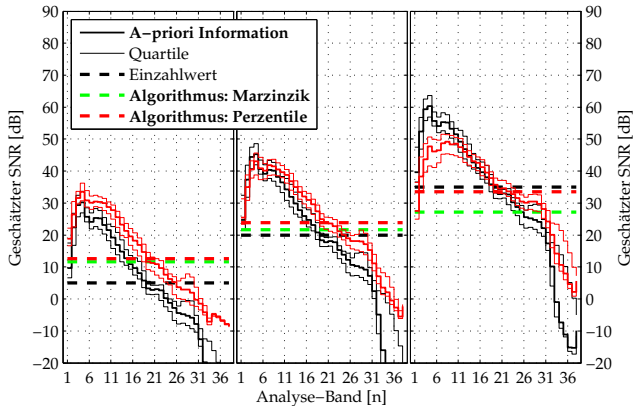



**Abbildung 2:** Exemplarische Abbildung der Mapping-Kurven von vier ausgewählten Sprechern, die sich im Anteil der Sprachpausen unterscheiden. Dies führt zu einer Varianz des Offset der Wertekorrektur.

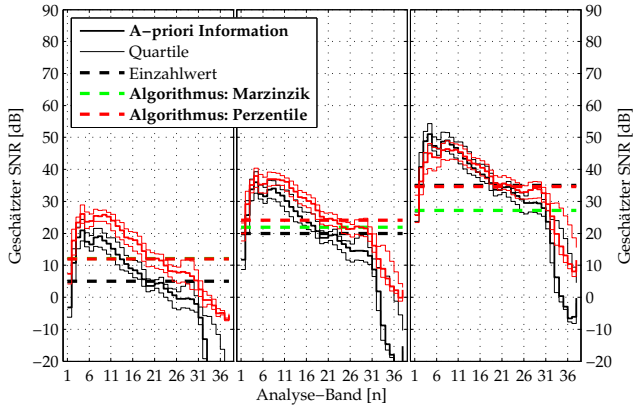
## Evaluation und Ergebnisse

Die Ergebnisse der Analyse der Sprachsignale wurden mit einem Schätzverfahren nach Marzinzik[2] verglichen, das auf Sprachpausendetektion (engl. *Voice Activity Detection*) basiert und für Sprache und stationäres weißes Rauschen hohe Übereinstimmungen erreichen kann. Die Abbildungen zeigen außerdem das jeweils aus den A-prioi-Informationen gewonnene wahre SNR-Profil.

Es zeigt sich für Sprachsignale (siehe Abbildung 3) eine hohe Übereinstimmung mit tatsächlichen SNR-Werten bei der Betrachtung in Bändern. Auch nach der Mittelung des SNR-Profils auf einen Einzahlwert erreicht die Schätzung eine hohe Genauigkeit. Jedoch tritt zu geringen wahren Signal-Rauschabständen eine Überschätzung des Algorithmus auf. Der Schätzfehler nimmt zu höheren SNR hin ab und der A-Priori-Wert wird sehr gut reproduziert.




(a) Rauschspektrum: weiß




(b) Rauschspektrum: rosa

**Abbildung 3:** Abbildung der Analyseergebnisse für Sprache mit additivem Rauschen im Vergleich zur A-priori Information. Der geschätzte SNR ist dabei gegenüber den Analysebändern aufgetragen. Gestrichelte Linien zeigen zudem gemittelte Einzahlwerte.

Beim Vergleich der Schätzung des weißen Rauschens mit rosa Rauschen (siehe Abbildung 3 (b)) wird ersichtlich, dass die Schätzung gegenüber einer spektralen Färbung des Rauschens robust ist und sowohl pro Band als auch gemittelt eine hohe Übereinstimmung mit den tatsächlichen Werten erreicht wird.

Bei Musik ist ausschließlich der Vergleich der Schätzung mit tatsächlichen Werten (siehe Abbildung 4) möglich, da ein Verfahren auf Basis der Sprachaktivität hier unzureichend funktioniert. Das ermittelte SNR-Profil zeigt nur mäßige Übereinstimmung mit der A-priori-Information. Besonders tieffrequente Bänder mit hohem wahren SNR weisen ein Sättigungsverhalten auf, das mit steigendem SNR auch in höheren Bändern aufzutreten scheint. Eine mögliche Ursache für die deutlich schlechteren Ergebnisse gegenüber der Schätzung bei Sprache ist die mit Hilfe von Sprachmaterial ermittelte Mapping-Kurve, da sich das Pausenverhalten von Sprache und Musik stark unterscheidet, bzw. bei Musik oft keinerlei Pause in den betrachteten 5s Blöcken und Frequenzbändern vorliegt.

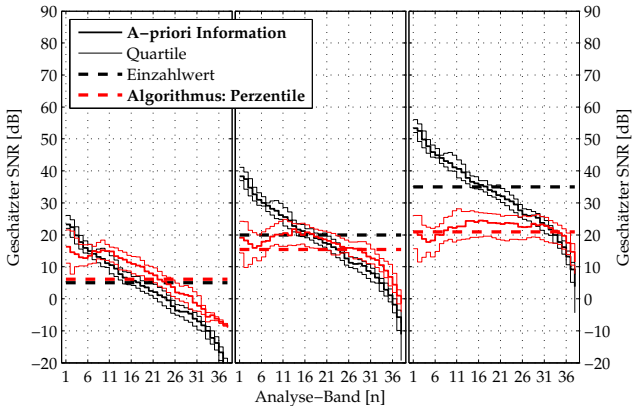



**Abbildung 4:** Abbildung der Analyseergebnisse für Musik mit additivem weißen Rauschen im Vergleich zur A-priori Information. Der geschätzte SNR ist dabei gegenüber den Analysebändern aufgetragen. Gestrichelte Linien zeigen zudem gemittelte Einzahlwerte.

## Fazit

Trotz Präzisionsverlust bei geringem SNR mit einhergehender Überschätzung bietet die Perzentilanalyse unabhängig der spektralen Farbe des Rauschens eine gute Reproduktion des frequenzabhängigen globalen Signal-Rauschabstands bei Sprache. Die Betrachtung des SNR-Profils liefert zudem qualitative Anhaltspunkte über die Färbung des Rauschens. Ursache für die Überschätzung ist möglicherweise das Sättigungsverhalten des Perzentilverhältnisses, das sich bereits im flachen Verlauf der Wertezuweisung zu geringem SNR zeigt.

Die Schätzung für verrauschte Musik ist noch unzureichend und liefert keine adäquaten Ergebnisse. Ursache ist das abweichende Perzentilverhältnis zwischen Rausch- und Musikleistung. Dies lässt vermuten, dass nach Ermittlung einer dedizierten Wertezuweisungskurve auch für Musik bessere Ergebnisse erzielt werden könnten.

## Literatur

[1] Bitzer, Jörg: *Mehrkanalige Geräuschunterdrückungssysteme – eine vergleichende Analyse*, Universität Bremen, Diss., 7 2001

[2] Marzinzik, Mark ; Kollmeier, Birger: Speech pause detection for noise spectrum estimation by tracking power envelope dynamics. In: *Speech and Audio Processing, IEEE Transactions on* 10 (2002), 2, Nr. 2, S. 109–118

Gefördert vom BMBF (Förderkennzeichen 03FH030PX2). Den Inhalt und die Erkenntnisse verantworten nur die Autoren.