

Kurzfassung der Dissertation von Timo Florian Gerkmann
"Statistical Analysis of Cepstral Coefficients and Applications in Speech Enhancement"
Lehrstuhl für Allgemeine Informationstechnik und Kommunikationsakustik

Von digitalen Kommunikationsgeräten, wie Hörhilfen oder Mobiltelefonen, werden häufig Sprachsignale erfasst, die durch Umgebungsgeräusche gestört sind. Mit Hilfe von Geräuschreduktionsalgorithmen kann das Verhältnis zwischen Sprach- und Geräuschsignalleistung (SNR) verbessert werden. Die Verbesserung des SNRs geht allerdings oft mit Prozessierungsartefakten oder Sprachsignalverzerrungen einher. Ziel dieser Arbeit ist es, ohne Einbußen in der Geräuschreduktion Artefakte und Sprachsignalverzerrungen zu reduzieren.

Bei den in dieser Arbeit verwendeten Geräuschreduktionsalgorithmen wird das gestörte Sprachsignal mit einer gleitenden diskreten Fouriertransformation (DFT) in den Spektralbereich transformiert, und dann in Abhängigkeit einer adaptiven Schätzung des a priori SNRs mit einer Gewichtungsfunktion multipliziert. Ist das a priori SNR lokal überschätzt, kommt es zu Ausreißern im prozessierten Sprachsignal, die häufig als tonale Artefakte wahrgenommen werden. Eine Unterschätzung führt hingegen zu Sprachsignalverzerrungen. Das Cepstrum bezeichnet die inverse diskrete Fouriertransformierte des logarithmierten Betragsquadrats einer spektralen Größe, und hat die Eigenschaft, dass spektrale Ausreißer einerseits und spektrale Sprachaktivität andererseits durch überwiegend disjunkten Mengen cepstraler Koeffizienten repräsentiert werden. Daher lassen sich, ohne wesentliche Zunahme der Sprachsignalverzerrungen, durch eine selektive Glättung im Cepstralbereich spektrale Ausreißer effektiv reduzieren.

In dieser Arbeit werden die statistischen Eigenschaften cepstraler Koeffizienten analysiert. Zunächst erfolgt die Herleitung von Gleichungen für den Mittelwert und die Varianz cepstraler Koeffizienten und logarithmierter spektraler Betragsquadrate. Es wird gezeigt, dass eine χ^2 -verteilte spektrale Zufallsgröße nach einer cepstralen Glättung noch annähernd χ^2 -verteilt ist, allerdings eine erhöhte Zahl Freiheitsgrade aufweist. Es wird eine Gleichung hergeleitet, welche die Bestimmung der Freiheitsgrade nach einer cepstralen Glättung in Abhängigkeit der Glättungsparameter erlaubt. Mit Hilfe dieser Ergebnisse wird nachgewiesen, dass eine erwartungstreue Glättung im Cepstralbereich zu einem systematischen Fehler im Spektralbereich führt. Wenn spektrale Größen wie das a priori SNR im Cepstralbereich geglättet werden, führt dieser systematische Fehler zu einer Unterschätzung der spektralen Gewichtungsfunktion und kann somit Sprachsignalverzerrungen hervorrufen. In dieser Arbeit wird der systematische Fehler in Abhängigkeit der verwendeten Glättungsparameter analytisch beschrieben, und kann somit korrigiert werden. Dadurch wird eine cepstrale Glättung möglich, die auch im Spektralbereich erwartungstreu ist.

Zur Bestimmung der Untermenge sprachrelevanter Cepstralkoeffizienten ist eine Schätzung der Sprachgrundfrequenz nötig. Es wird hergeleitet, dass hierfür die Suche des Maximums der Cepstralkoeffizienten, welche die spektrale Feinstruktur darstellen, optimal im Maximum-Likelihood Sinn ist. Sind mehrere Mikrofone vorhanden, so erhält man die optimale Lösung, wenn die Cepstren der Mikrofonensignale vor der Maximumssuche addiert werden. Weiterhin kann die Korrelation der Sprachgrundfrequenz zeitlich aufeinanderfolgender Segmente für eine robustere Schätzung im Sinne eines Maximum-A-Posteriori Schätzers genutzt werden.

Es wird belegt, dass eine zeitliche Glättung des Cepstrums spektraler Größen, wie der a priori SNR Schätzung, spektrale Ausreißer reduziert und unter Verwendung des hergeleiteten Korrekturterms im Vergleich zu herkömmlichen Verfahren sowohl zu einer Erhöhung des Ausgangs-SNRs als auch zu weniger Sprachsignalverzerrungen führt. Ähnliche Ergebnisse werden erzielt wenn cepstrale Koeffizienten des prozessierten Signals durch Koeffizienten des unprozessierten Signals ersetzt werden. Der Vorteil der cepstralen Ersetzung im Gegensatz zu einer zeitlichen cepstralen Glättung ist, dass eine zeitliche Verzerrung des prozessierten Signals ausbleibt. Dadurch wird insbesondere bei nichtstationären Geräuschquellen, wie konkurrierenden Sprechern, die Natürlichkeit des prozessierten Signals weiter erhöht. Der Vorteil der zeitlichen cepstralen Glättung ist hingegen ein geringerer Rechenaufwand. Zudem wird gezeigt, dass der Rechenaufwand durch eine Modifikation der Spektraltransformationen, die zur Berechnung des Cepstrums verwendet werden, deutlich reduziert werden kann, während die Qualität des Ausgangssignals annähernd unverändert bleibt.

Abschließend wird die Schätzung der a posteriori Sprachanwesenheitswahrscheinlichkeit (SPP) in jedem spektralen Koeffizienten behandelt. Im Gegensatz zu konkurrierenden Verfahren werden a priori SNR und a priori SPP nicht adaptiert, sondern vorab bestimmt, was zu einer Entkopplung der a posteriori SPP Schätzung und der Schätzung der spektralen Gewichtungsfunktion führt. Der vorgeschlagene Algorithmus führt zu weniger Schätzfehlern als konkurrierende Verfahren. Dies wird besonders dann deutlich, wenn der SPP Schätzer mit einer zeitlichen Glättung des Cepstrums, dem hergeleiteten Korrekturterm und dem vorgeschlagenen Verfahren zur Bestimmung der Freiheitsgrade kombiniert wird.