# EIN EFFEKTIVER ALGORITHMUS ZUR KOMBINIERTEN ECHOKOMPENSATION UND GERÄUSCHREDUKTION IN FREISPRECHANWENDUNGEN

*Rico Petrick*, Christian Gruber*, Dr. Michael Fenske***

**voice INTER connect GmbH*
***Funkwerk Dabendorf GmbH*
*[petrick, gruber]@voiceinterconnect.de*
*michael.fenske@fwd-online.de*

**Abstract:** Freisprecheinrichtungen, wie sie vor allem beim Telefonieren im Auto eingesetzt werden, haben gegenüber dem klassischen Telefon den Nachteil, dass das Sprachsignal wesentlich stärker von Störgeräuschen überlagert ist. Dazu zählen sowohl Umgebungsgeräusche als auch Rückkopplungen des Sprachsignals des anderen Gesprächsteilnehmers über den Lautsprecher in das Mikrofon. Um diese Effekte zu verringern und damit die Sprachqualität zu verbessern werden Geräuschreduktions- und Echokompensationsfilter eingesetzt. Viele Ansätze zur Kombination beider Filter implementieren diese getrennt voneinander. Dies hängt vor allem mit der Funktionsweise der Algorithmen zusammen, die sich unabhängig voneinander entwickelt haben. Echokompensationsfilter arbeiten hauptsächlich im Zeitbereich. Die meisten Algorithmen zur Geräuschreduktion arbeiten dagegen im Frequenzbereich. Dieser Artikel zeigt einen Ansatz, wie sich Echokompensation und Geräuschreduktion im Frequenzbereich effektiv miteinander verbinden lassen. Damit lässt sich der Rechenaufwand verringern, ohne dass die Funktionsweise von Echokompensation und Geräuschunterdrückung beeinträchtigt wird. Praktische Ergebnisse werden anhand eines Demonstrators gezeigt.

## 1 Einleitung

Beim Telefonieren mit einer Freisprecheinrichtung im Automobil wird die Qualität des vom Mikrofon aufgenommenen Sprachsignals *s(t)* von zwei unterschiedlichen Klassen von Störquellen gemindert (vgl.: Abbildung 1).

- **systemfremde Störgeräusche:** Dies ist die Summe aller Umgebungsgeräusche *n(t)*, wie z.B. Fahrtwind oder Motorengeräusche welche vorrangig von außen in den Fahrzeuginnenraum gelangen. Da sie aus Sicht der Freisprecheinrichtung von externer Quelle stammen, kann man auch von systemfremden Störgeräuschen sprechen.
- **systemeigene Störgeräusche:** Die zweite Quelle ist das in das Mikrofon eingekoppelte Lautsprechersignal *y(t).* Dies können Signale der Car-HiFi Anlage oder das Telefonsignal des fernen Teilnehmers sein. Durch die akustische Rückkopplung hört sich der ferne Sprecher selbst (akustisches Echo). Da diese Signalquelle aus Sicht der Freisprecheinrichtung bekannt ist und zum System selbst gehört, kann man von systemeigenen Störgeräuschen sprechen.

Beide Störgeräuschklassen werden aufgrund ihres unterschiedlichen Charakters mit verschiedenen Methoden unterdrückt. Systemfremde Geräusche im Kfz sind meist quasistationär, d.h. für das System hinreichend langsam veränderlich. Dafür existieren verschiedene Ansätze zur automatischen Störgeräuschunterdrückung (Automatic Noise Reduction ANR). Für systemeigene Störgeräusche gibt es verschiedene Ansätze zur akustischen Echokompensation (Acoustic Echo Cancellation AEC).

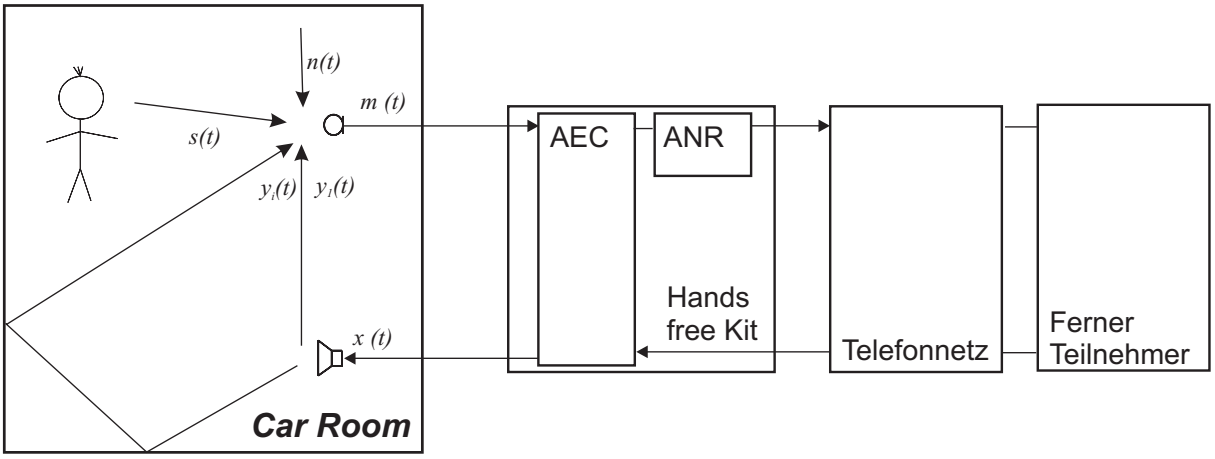


**Abbildung 1: Störgeräuschquellen beim Freisprechen**

Es wurde eine AEC/ANR-Einheit für Freisprecheinrichtungen entwickelt, die für beide Teilaufgaben geeignete Algorithmen kombiniert. Diese wurde für den Automobileinsatz optimiert und auf einer „embedded“ Plattform, dem ADSP2153x Blackfin DSP von Analog Devices, implementiert.

# 2 Automatische Geräuschreduktion (ANR)

## 2.1 Spektrale Subtraktion

Für die Lösung dieser Aufgabe wurde der Ansatz der Spektralen Subtraktion [3] gewählt. Zwei oder mehrkanalige Verfahren sind aus Kostengründen zunächst auszuschließen.

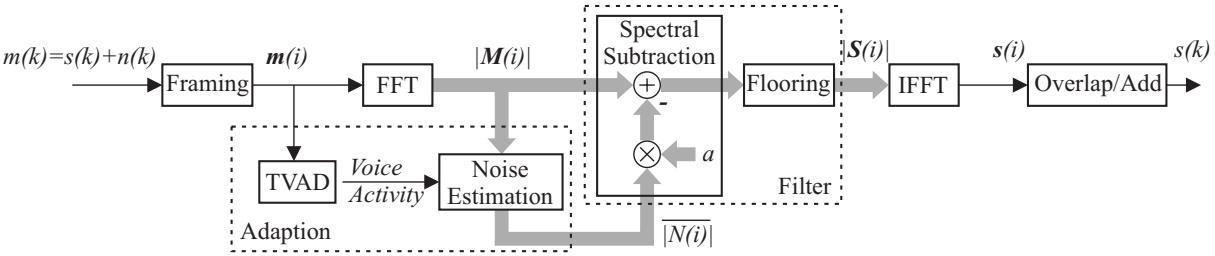


**Abbildung 2: Blockschaltbild[1] der ANR mit Spektraler Subtraktion**

Abbildung 2 zeigt die Realisierung der Geräuschunterdrückung mittels Spektraler Subtraktion. Das vom Mikrofon aufgenommene digitalisierte Signal *m(k)* wird zunächst in Blöcke (Frames) geteilt, welche in Echtzeit bearbeitet werden und danach mittels Overlap-and-Add wieder zu einem kontinuierlichen Zeitsignal zusammengesetzt werden.

Das Mikrofonsignal *m(k)* entsteht aus der Überlagerung des Sprachsignals *s(k)* und der Summe der Störgeräusche *n(k):*

$$m(k) = s(k) + n(k)$$

[1]Fette Buchstaben symbolisieren einen Vektor (Blockzeitsignal z.B. ***m(i)*** oder Blockspektrum z.B. ***M(i)*** mit dem Blockindex *i* einer Blockfolge; die Indizes der Vektorkomponenten (Zeit, Frequenz) werden nicht dargestellt). Schlanke Buchstaben symbolisieren Abtastwertfolgen.

Unter der idealen Annahme, dass *s(k)* und *n(k)* nicht miteinander korreliert sind, addieren sich die Leistungsdichtespektren der beiden Signale. Bei blockweiser Verarbeitung werden die Kurzzeitleistungsspektren des Mikrofonsignals $|M(\Omega)|^2$, des Sprachsignals $|S(\Omega)|^2$ und des Geräuschsignals $|N(\Omega)|^2$ benutzt. Durch Umstellen erhält man die Spektrale Subtraktion:

$$|S(\Omega)|^2 = |M(\Omega)|^2 - \alpha \cdot |N(\Omega)|^2$$

$$1{,}2 \leq \alpha \leq 3$$

$\alpha$ ist ein wichtender Subtraktionskoeffizient der die Stärke der Geräuschunterdrückung einstellt. Nach der Spektralen Subtraktion folgt das Flooring, welches zum einen das Vorkommen von negativen Elementen von $|S(\Omega)|^2$ verhindert. Zum anderen wird über einen Flooringfaktor $\beta$ ein Restrauschen eingestellt, welches eine subjektiv angenehme Sprachqualität hervorrufen soll.

$$|S(\Omega)|^2 = \begin{cases} |S(\Omega)|^2 & ; \quad |S(\Omega)|^2 > \beta |\cdot N(\Omega)|^2 \\ \beta \cdot |N(\Omega)|^2 & ; \quad |S(\Omega)|^2 \leq \beta |\cdot N(\Omega)|^2 \end{cases}$$

$$0 \leq \beta \leq 0{,}5$$

Für die Dimensionierung von $\alpha$ und $\beta$ gibt es verschiedene Varianten. Sie können für das gesamte Frequenzband oder auch in Subbändern, sowohl statisch als auch dynamisch eingestellt werden. Eine dynamische Einstellung erfolgt in Abhängigkeit vom SNR.

## 2.2 Adaptives Verhalten – Voice Activity Detector (VAD)

Das spektral subtrahierte Störgeräuschspektrum ist eine Unbekannte. Zur Ermittlung des Störgeräuschspektrums wird in Sprachpausen $|N(\Omega)|^2$ geschätzt. Ein Voice Activity Detector im Zeitbereich (TVAD) nach [11] scannt dabei das Blocksignal ***m(i)*** und markiert den Block als Sprache oder Pause. Zur Schätzung des Störgeräusches werden nur Pausenblöcke herangezogen. Die Schätzung erfolgt in Form einer gleitenden Mittelung. Die gleitende Mittelung kann in ihrem Verhalten als Tiefpass aufgefasst werden.

$$|N(\Omega,i)| = \rho \cdot |N(\Omega,i-1)| + (1-\rho) \cdot |M(\Omega,i)|$$

$$0.7 \leq \rho < 1$$

Der Parameter $\rho$ wird aufgrund seiner Funktion als Glättungsfaktor bezeichnet. Er bestimmt die Trägheit der Adaption.

Der Tiefpass entfernt schnelle Schwankungen in der Folge der Geräuschspektren, langsame Schwankungen werden erfasst. Dadurch wird eine Adaption der Geräuschunterdrückung an Änderungen der Geräuscheigenschaften (z.B. bei Beschleunigung eines Fahrzeugs) erreicht. Instationäre Störgeräusche wie z.B. Türenklappen können mit diesem Verfahren nicht berücksichtigt werden.

## 2.3 Sprachqualität und Reduktion von Musical Tones

Der Subtraktionskoeffizient $\alpha$ ist so einzustellen, dass das Geräusch hinreichend subtrahiert wird. Durch die starke Schwankung der Kurzzeitgeräuschspektren um ihren Mittelwert kommt es vor, dass teilweise zu viel, bzw. zu wenig subtrahiert wird. Eine zu starke

Subtraktion verzerrt das Nutzsignal, eine zu schwache lässt vereinzelte Frequenzspitzen des Rauschens übrig, die zu Musical Tones[2] führen. Wird α mit Werten nahe 2 besetzt, erhält man für die meisten SNR eine günstige Konstellation von Sprachqualität, Geräuschreduktion und Musical Tones. Verbleibende Musical Tones werden mit einem Postfilter[3] reduziert.

# 3 Akustische Echokompensation (AEC)

## 3.1 Lösungsansätze

Das vom Lautsprecher ausgesendete Signal $x(k)$ wird durch das als linear angenommene System Lautsprecher-Raum-Mikrofon (LRM-System) verändert. Das Echosignal $y(k)$ besteht sowohl aus dem direkt eingekoppelten Schallsignal als auch aus dem Signal des diffusen Schallfeldes. Der Ansatz bei der AEC ist ein künstliches (digitales) System $\boldsymbol{W}(i)$, welches das Verhalten des LRM-Systems nachbildet. Das Lautsprechersignal wird in dieses System eingeführt und das am Ausgang abgegriffene Signal $\hat{y}(k)$ kann vom Mikrofonsignal $m(k)$ subtrahiert werden. Da das LRM-System zeitveränderlich ist, wird ein adaptives Filter benutzt, welches das Übertragungsverhalten des Raumes schätzt.

Bekannte Ansätze für das adaptive Filter sind:

**Zeitbereichsverfahren:**

- LMS-Algorithmus: Least Mean Square Verfahren (am weitesten verbreitet) und Zusatzmaßnahmen wie Normierung (NLMS) [4], adaptive Schrittweitensteuerung [3], Dekorrelationsfilter [6], Echoshapingfilter [3], Center Clipper [3]
- RLS-Algorithmus [4]: Recursive Least Square
- AP-Algorithmus [4]: Affine Projektion
- FTF-Algorithmus [4]: Fast Transversal Filter
- FN-Algorithmus [4]: Fast Newton

**Frequenzbereichsverfahren (Frequency Domain Adaptive Filters FDAF):**

- FBLMS-Algorithmus [5]: Frequency Domain Block LMS

Die Algorithmen unterscheiden sich in Adaptionsverhalten und -geschwindigkeit, numerischer Komplexität und benötigter Rechenleistung, erreichbarem ERLE[4]-Wert, Robustheit gegenüber verschiedenen akustischen Umgebungen sowie der Qualität des gereinigten Sprachsignals.

## 3.2 Der Echokompensationsalgorithmus

Frequenzbereichs-Block-LMS-Algorithmen vereinen die Vorteile verschiedener Gruppen von Algorithmen. Durch die blockweise Verarbeitung verringert sich die numerische Komplexität gegenüber Zeitbereichsverfahren. Der Rechenaufwand pro Iterationsschritt ist zwar größer, es sind jedoch weniger Iterationsschritte pro Zeitintervall nötig [4]. Durch die Verarbeitung im Frequenzbereich lässt sich eine frequenzselektive Schrittweitensteuerung implementieren, welche die Konvergenzgeschwindigkeit entscheidend vergrößert und unabhängig gegenüber dem Sprachsignal des nahen Gesprächsteilnehmers macht.

---

[2] Musical Tones: Einzelne kurze Sinustöne unterschiedlicher Frequenz

[3] Postfilter z.B. nach einem Medianprinzip [10].

[4] ERLE (Echo Return Loss Enhancement [3]) in dB beschreibt die Stärke der Echounterdrückung.

Abbildung 3 zeigt den Aufbau des verwendeten FBLMS-Algorithmus. Das Lautsprechersignal *x(k)* und das Mikrofonsignal *m(k)* werden zu Blöcken ***x(i)*** und ***m(i)*** zusammengefasst (Framing). Die Berechnung des Schätzsignals *ŷ(k)* erfolgt durch schnelle Faltung [2] (Multiplikation im Frequenzbereich) mit dem adaptiven Filter ***W(i)***. Die Filterkoeffizienten ***W(i)*** liegen im Frequenzbereich vor. Die Berechnung des Gradientenvektors Δ***W(i)*** zur Anpassung von ***W(i)*** erfolgt ebenfalls im Frequenzbereich durch eine Korrelationsanalyse.

$$\text{Zeitbereich}: \quad \Delta \boldsymbol{w}(i) = \sum_l \boldsymbol{x}(iL-l) \cdot q(iL-l) \quad L \ldots \text{Framelänge}$$

$$\text{Frequenzbereich}: \quad \Delta \boldsymbol{W}(i) = \boldsymbol{X}(i)^* \cdot \boldsymbol{Q}(i)$$

Die zusätzliche IFFT mit anschließendem Windowing und erneuter FFT ist notwendig, um Effekte der zirkularen Korrelation zu vermeiden [2]. Die Anpassung von ***W(i)*** erfolgt durch

$$\boldsymbol{W}(i+1) = \boldsymbol{W}(i) + \boldsymbol{a}(i) \cdot \Delta \boldsymbol{W}(i)$$

Die Komponenten des Schrittweitenvektors ***a(i)*** stellen für jede Frequenz einen eigenen Schrittweitenfaktor dar. Die Schrittweitensteuerung [5] erfolgt adaptiv, die Schrittweitenfaktoren hängen vom Blockindex *i* und von der Frequenz ab. Durch Overlap-and-Add wird aus dem Blocksignal ***q(i)*** das Ausgangszeitsignal der Echokompensation *q(k)* wieder hergestellt.

Um das Problem von großen Blocklängen bei langen Impulsantworten und damit von zu großen Verzögerungszeiten zu vermeiden, wird ein großer in mehrere kleine Blöcke zerlegt, welche nacheinander abgearbeitet werden (im Blockschaltbild nicht dargestellt).

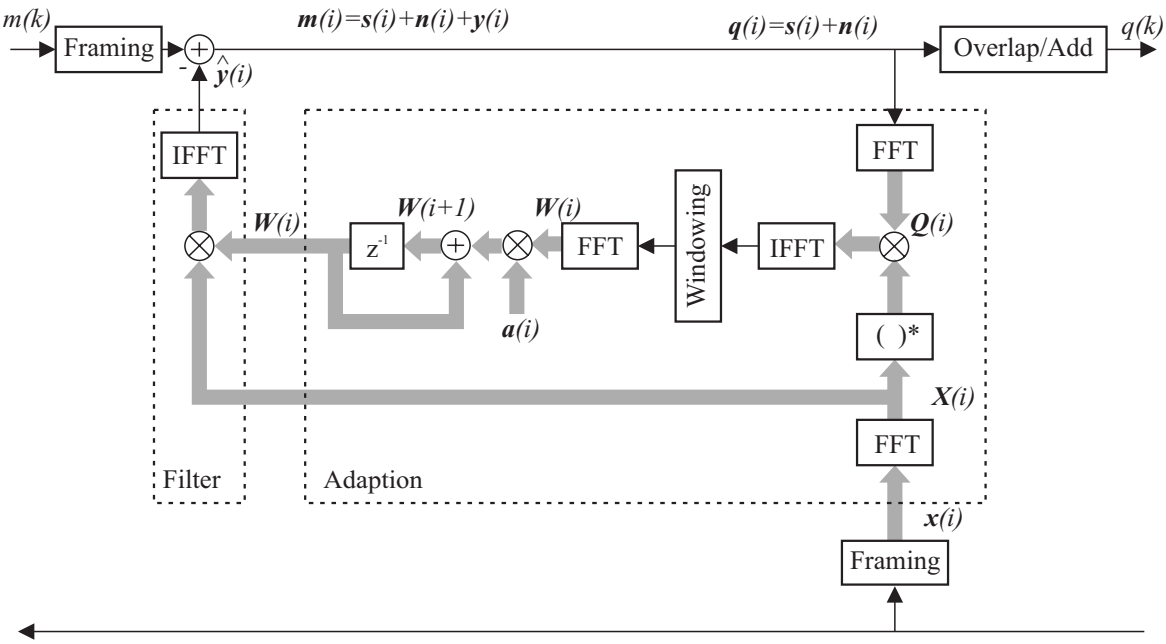


**Abbildung 3: Blockschaltbild der AEC mit dem FBLMS-Algorithmus**

## 4 Kombination von ANR und AEC Algorithmen

AEC und ANR sind nacheinander geschaltet. Durch geschickte Verkoppelung von AEC und ANR im Frequenzbereich lassen sich verschiedene Rechenoperationen einsparen. Es entfallen die Ermittlung des Signals *q(k)* durch Overlap-and-Add nach der AEC sowie Framing und FFT vor der ANR.

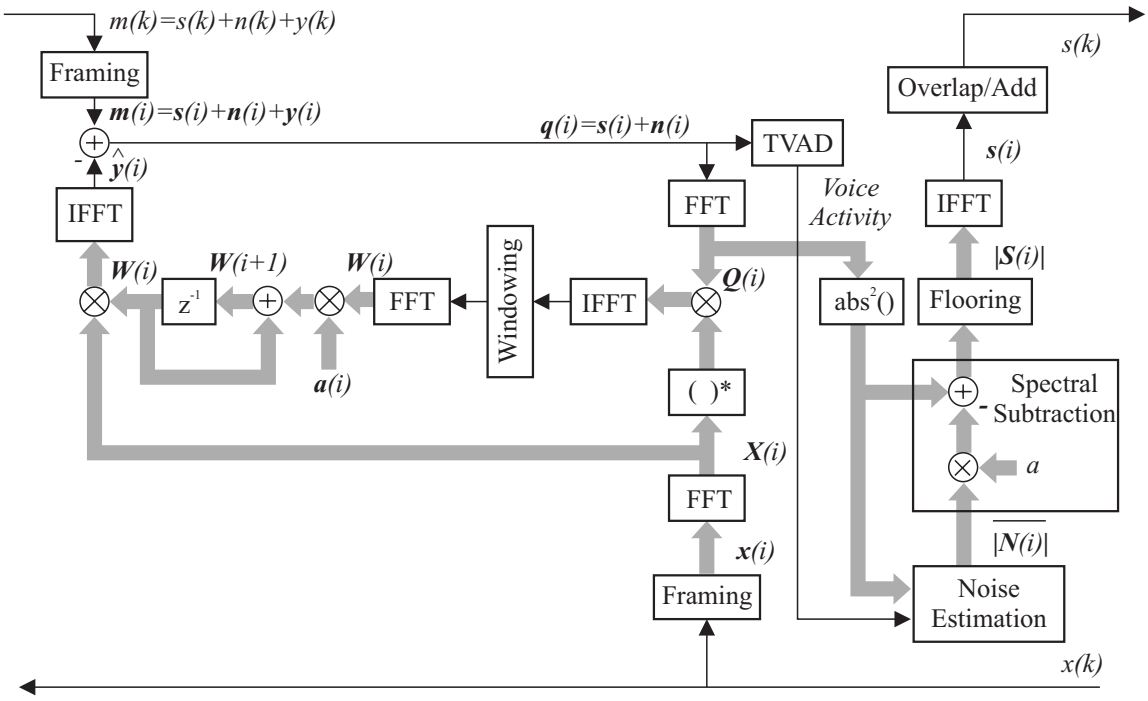


**Abbildung 4: Blockschaltbild der kombinierten AEC und ANR**

# 5 Evaluierung

Zur Evaluierung des Systems wurden in Zusammenarbeit mit der Funkwerk Dabendorf GmbH Experimente in einer realen Testumgebung (original ausgestattete LKW-Fahrerkabine) durchgeführt. Die Fahrerkabine war mit 2 Personen besetzt und die Türen waren geschlossen (Zustand bewegtes Fahrzeug). Die Klimaanlage wurde mit 3/4 der maximalen Kühlleistung betrieben. Es wurden 3 Mikrofone an unterschiedlichen Positionen in der Umgebung des Fahrers montiert:

- Pos. 1: Mikrofon mit Nierencharakteristik, angebracht über der Windschutzscheibe an der oberen Verblendung halbrechts vom Fahrer, Abstand ca. 60 cm
- Pos. 2: Mikrofon mit Nierencharakteristik, angebracht auf der Mittelkonsole vor der Frontscheibe, Abstand ca. 100 cm
- Pos. 3: Mikrofon mit Nierencharakteristik angebracht an der oberen Verblendung in der Mitte der Windschutzscheibe, Abstand ca. 80 cm.

Die Mikrofonsignale wurden getrennt an die weiterverarbeitenden Geräte geführt: Pos. 1 an die AEC/ANR Einheit der voice INTER connect GmbH (VIC-ANR/AEC). Das unverarbeitete Signal und das gereinigte Signal der VIC-AEC/ANR wurde digital in 2 Spuren eines Stereo-Wave-Files (16kHz/Lin16) auf einem mobilen PC aufgezeichnet, ebenso wurde das Rohsignal aus Pos 2 und Pos 3 analog aufgezeichnet. Die Aufnahme des Mikrofonsignals und des gereinigten Signals erfolgten je Szenario für 60 s. Der Sprecher auf dem Fahrersitz in der normalen Fahrposition las einen vorgegebenen Text vor.

Zur Evaluierung wurden verschiedene akustische Szenarien getestet. Von den getesteten Szenarien wurden für die Auswertung drei Störgeräuschfälle betrachtet:

- Motor im Stillstand
- Motor 800 Umdrehungen pro min (mittlere Drehzahl)
- Motor 2000 Umdrehungen pro min (nahe Maximaldrehzahl)

Zum Test der Echounterdrückung wurde eine Lautsprechergruppe auf der Mittelablage zwischen den Sitzen positioniert und in ca. 145 cm Abstand auf das Mikrofon Pos1. gerichtet. Es wurde ohne und mit Echosignal getestet. Als Echosignal wurde ein modernes Musikstück mit großer Lautstärke benutzt, um die Grenzen des Systems unter Extrembedingungen zu testen. Ein Musikstück ist stark korreliert, was für LMS-Algorithmen ungünstige Adaptionsbedingungen hervorruft. Die Lautstärke wurde so angepasst, dass Sprecher- und Echosignal am Mikrofon in etwa den gleichen Signalpegel aufweisen. Tabelle 1 zeigt die Evaluationsergebnisse. Da das Sprechersignal *s(t)* nicht isoliert vorliegt, wurden die SNRs wie folgt berechnet:

$$SNR = 10 \cdot \log_{10} \frac{S}{N} = 10 \cdot \log_{10} \frac{M_s - N}{N} = 10 \cdot \log_{10} \frac{m_{s,eff}^{\ 2} - n_{eff}^{\ 2}}{n_{eff}^{\ 2}}$$

**Tabelle 1: SNR vor und nach AEC/ANR Einheit, Dämpfung der Störgeräusche (NR&ERLE)**

| Versuch: | Akustisches Szenario | SNR vor AEC/ANR | SNR nach AEC/ANR | NR&ERLE |
|---|---|---|---|---|
| 1 | kein Echo / Motor 800 $Umin^{-1}$ | 13.5 dB | 26.7 dB | 15.3 dB |
| 2 | kein Echo / Motor 2000 $Umin^{-1}$ | 2.8 dB | 21.0 dB | 20.2 dB |
| 3 | Musikecho / Motor aus | 7.7 dB | 27.9 dB | 21.7 dB |
| 4 | Musikecho / Motor 800 $Umin^{-1}$ | 3.9 dB | 21.4 dB | 22.0 dB |
| 5 | Musikecho / Motor 2000 $Umin^{-1}$ | 2.2 dB | 17.9 dB | 17.6 dB |

Dabei wurde die Leistung des Rauschsignals *N* aus dem Effektivwert des Mikrofonsignals in Sprechpausen $m_{p,eff} = n_{eff}$ ermittelt. *N* beinhaltet demnach die Summe aller Geräusche, d.h. sowohl die Motorgeräusche als auch das eingespielte Echo. Die Leistung des Sprachsignals wurde aus dem Effektivwert des Mikrofonsignals während Sprachaktivität $m_{s,eff}$ ermittelt.

ANR: Die Messungen 1 und 2 zeigen die alleinige Funktion der ANR, da kein Echosignal eingespielt ist. Das LKW-Motorgeräusch wird bis zu 20 dB gedämpft. Die gereinigten Signale sind gut verständlich.

AEC: Messung 3 zeigt die alleinige Funktion der AEC, da kein Motorgeräusch vorhanden ist (Stillstand des Motors). Die Echounterdrückung beträgt ca. 22 dB. Die gereinigten Signale sind sehr gut verständlich. Das Restecho ist kaum hörbar bei unverzerrtem Sprachsignal.

AEC/ANR: Bei den Messungen 4 und 5 liegt sowohl ein Motorgeräusch als auch ein Echo an. Sowohl subjektiver Höreindruck als auch die gemessenen SNRs zeigen die Funktionsfähigkeit auch unter sehr ungünstigen akustischen Bedingungen. Die Dämpfung der Geräuschunterdrückung sowie ERLE (Spalte NR&ERLE) können nicht einfach aufaddiert werden, da bei Vorhandensein von Störgeräuschen die AEC nur Echoanteile kompensiert, die stärker als der Störgeräuschpegel sind. Kleinere Echoanteile verschwinden im Geräusch und werden teilweise von der ANR beseitigt.

## 6 Ausblick

Die Ergebnisse der Experimente weisen die Funktionsfähigkeit der vorgestellten AEC/ANR-Einheit nach. Aus den durchgeführten Messungen lassen sich qualitative Aussagen machen, für den vollständigen Nachweis von Standardkonformität waren sie auch nicht gedacht . Dennoch ist ein solcher Test in einer realen einsatznahen Testumgebung wichtig, da viele Algorithmen bisher nur in Simulationsumgebungen z.B. MATLAB evaluiert wurden und somit die Effekte des praktischen Einsatzes unberücksichtigt blieben.

Um verschiedene Systeme miteinander zu vergleichen, ist es nötig, ein einheitliches, eventuell genormtes Messverfahren zu verwenden. Dazu ist ein einheitlicher geometrischer Aufbau mit Kfz-Raum, Lautsprecher, Mikrofon, Sprechersignalquelle sowie verschiedenen Störsignalquellen vorgeschrieben. Um Reproduzierbarkeit sicherzustellen, müssen unterschiedliche Pegelkombinationen von standardisierten Teilgeräuschen *s(t), n(t), x(t)* eingespielt werden. Die Aufnahme sowie Auswertung kann automatisiert werden. Mögliche Auswertekriterien sind Adaptionszeiten, erreichbare Störgeräuschunterdrückung, erreichbarer ERLE-Wert und Systemdelay. Nicht erfasst werden dabei die Flächenstrahler, die im bewegten NKW wesentlich das Hintergrundgeräusch ausprägen.

Ein sehr wesentlicher Faktor der Akzeptanz beim Anwender ist der subjektive Höreindruck. Aus unserer Sicht ist das das wichtigste Kriterium zur Bewertung von AEC oder ANR-Algorithmen, da oft der angenehme Höreindruck über die Akzeptanz beim Kunden entscheidet. Hierzu zählen ein möglichst unverzerrtes Sprachsignal mit dem Erhalt der Persönlichkeit des Sprechers, seiner individuellen Sprachmelodie und seiner Erkennbarkeit am fernen Ende, kaum hörbare Musical Tones, akzeptables Restrauschen, akzeptables Restecho sowie eventuell ein Comfort Noise. Natürlich ist die Erhöhung der Erkennungsrate von Sprachsteuerungen auch ein wichtiges Kriterium, zu dem in einer weiterführenden Arbeit detailliert Messergebnisse vorgestellt werden.

Um Systeme anhand von subjektiven Kriterien zu bewerten, ist die Bestimmung eines MOS[5]-Wertes durch eine möglichst große Anzahl an Probanden nötig. Solche Tests sind jedoch sehr aufwändig. Derzeit erfolgt eine Evaluierung der aufgenommenen Daten im Paarvergleichstest mit zahlreichen Probanden unter Verwendung von WiGE-test, einem eigens für subjektive Sprachdatentests entwickelten grafischen Tool [12].

## Literatur


[1] Petrick R.: “Maßnahmen zur Erhöhung der Robustheit eines schnellen Kommando-worterkenners”, Diplomarbeit, HTW Dresden, 2001

[2] Oppenheim/Schafer: “Zeitdiskrete Signalverarbeitung”, Oldenbourg-Verlag, 1995

[3] Vary, Heute, Hess: „Digitale Sprachsignalverarbeitung“, Teubner-Verlag, 1998

[4] Haykin, S.: “Adaptive Filter Theory”, Prentice Hall, 2002

[5] Farhang-Boroujeny, B.: “Adaptive Filters – Theory and Applications”, John Wiley & Sons, 1998

[6] Breining, Dreiseitel, Hänsler, Mader, Nitsch, Puder, Schertler, Schmidt, Tilp: “Acoustic Echo Control”, IEEE Signal Processing Magazine, 1999

[7] Gustafsson S., Jax P., Kamphausen A., Vary P.: “A Postfilter for Echo and Noise Reduction Avoiding the Problem of Musical Tones”, ICASSP, 1999

[8] Al-Naimi K., Sturt C., Kondoz A.: “A Robust Noise and Echo Canceller”, Eurospeech, 2003

[9] Enzner G.: “Hands-Free Communication: A Unified Concept of Acoustic Echo Cancellation and Residual Echo Suppression”, DAGA, 2003

[10] Goh Z., Tan K.-C., Tan B. T. G. “Postprocessing Method for Suppressing Musical Noise Generated by Spectral Subtraction” IEEE Transactions on Speech and Audio Processing,. Vol. 6, No. 3, MAY 1998 pp. 287-292

[11] Richter T., Petrick R., Hirschfeld D.: “Robuste Phrasendetektion durch zweistufige Sprach/Pause- Detektion“, 14. Konf. Elektron. Sprachsignalverarbeitung Karlsruhe 2003, vol. 28, pp. 46-53

[12] Koloska, Uwe: „Integrierte multimodale Sprachdatenverwaltung mit dem WiGE-System“, erscheint in: 15. Konf. Elektron. Sprachsignalverarbeitung Cottbus 2004.


[5] MOS - Mean Opinion Score