

Klangqualitätsvorhersage von Mehrkanal-Störgeräuschreduktionsverfahren in Personenkraftwagen

Rainer Huber

HörTech gGmbH - Kompetenzzentrum für Hörgerätesystemtechnik, D-26129 Oldenburg

E-Mail: Rainer.Huber@HoerTech.de

Einleitung

Zur Beurteilung der Leistungsfähigkeit von Störgeräuschreduktionsverfahren für Sprache werden oft recht einfache, technische Maße, wie z.B. die Verbesserung des Signal-zu-Rausch-Verhältnisses (SNR) verwendet. Insbesondere bei einkanaligen Verfahren wird mit zunehmender Geräuschabsenkung i.A. aber auch das Sprachsignal immer stärker verzerrt. Zur empfundenen Gesamtqualität tragen beide Effekte gegensätzlich bei. Die bloße Betrachtung der SNR-Verbesserung ist in solchen Fällen zur Abschätzung der Gesamtklangqualität offensichtlich nicht ausreichend.

In diesem Beitrag wird gezeigt, dass darüber hinaus auch bei nahezu verzerrungsfreier Verarbeitung durch mehrkanalige Störgeräuschreduktionsverfahren (hier: Beamformer) die Korrelation zwischen subjektiv beurteilter Gesamtqualität und gemessener SNR-Verbesserung unzufriedenstellend sein kann. Dagegen ergeben sich im betrachteten Fall z.T. sehr viel bessere Vorhersageergebnisse durch stärker gehörorientierte objektive Qualitätsmaße, die somit als Alternative bzw. Ergänzung zu aufwändigen und kostspieligen Hörtests in Betracht gezogen werden können.

Testdatenbasis

Die mögliche Eignung objektiver Qualitätsmaße zur Vorhersage subjektiver Gesamtqualitätsbewertungen wurde anhand einer Datenbasis subjektiv bewerteter Testsignale untersucht, die K. Tontch 2002 im Rahmen seiner Dr.-Arbeit generiert hat und dem Autor freundlicherweise zur Verfügung gestellt wurde. Tontch untersuchte verschiedene Arten von Beamformern für Freisprecheinrichtungen in Pkw [1]. Einige Beamformer ließ er durch formale Hörtests u.a. hinsichtlich der empfundenen Gesamtqualität evaluieren.

Die evaluierten Beamformer basierten z.T. auf 2D linearen, z.T. auf halbkreisförmigen Mikrofonarrays, deren Ausgänge verzögert aufsummiert wurden (Delay-and-Sum-Beamformer). Variiert wurden die Abstände zum Sprecher, die Kombination einiger Mikrofone zu Sub-Arrays, sowie die Art der Sprecherlokalisierung (automatisch vs. vorgegeben). (Für genauere Spezifikationen s. [1].)

Die Testsignale für die Hörtests bestanden aus einem 2s langen Satz, der von einem männlichen Sprecher und einer weiblichen Sprecherin in störgeräusch- und reflexionsarmer Umgebung aufgesprochen und später durch einen Head-and-Torso-Simulator mit künstlichem Mund (HEAD Acoustics) im Innern der Fahrgastzelle wiedergeben wurde. Die Aufnahme und Verarbeitung dieser Signale durch die Beamformer erfolgte bei verschiedenen Fahrzeuggeschwindigkeiten (100, 130, 160 km/h). Vier

Testkonditionen (d.h. Kombinationen von Sprecher/in und Geschwindigkeit) wurden für die subjektiven Hörtests ausgewählt.

Je Kondition wurde ein vollständiger Paarvergleichstest der Ausgangssignale von acht Beamformer-Varianten plus dem einkanalig aufgenommenen, unverarbeiteten Signal durchgeführt. (Ausnahme: sechs Beamformer in einer 160 km/h-Kondition). Beurteilungskriterium war die Gesamtpräferenz. 10 normalhörende Versuchspersonen (6 weibl., 4 männl., Durchschnittsalter = 32 Jahre) nahmen an den Hörtests teil. Die unmittelbaren Ergebnisse der vier Paarvergleichstests wurden durch Anwendung der Methode nach Bradley, Terry und Luce (BTL) [2] auf eine Intervallskala transformiert, wobei der Nullpunkt der Skala willkürlich so gewählt war, dass die Qualitätswerte der unverarbeiteten Signale auf null abgebildet wurden. Dadurch konnten die Ergebnisse aller Paarvergleichstests kombiniert werden.

Objektive Qualitätsmaße

Neun objektive Qualitätsmaße bzw. Schätzmethode wurden auf die Testsignale angewandt und den subjektiven Qualitätsbewertungen gegenübergestellt: Gesamt-SNR, segmentweises SNR (segSNR), Itakura-Saito Distortion (ISD), Log-Likelihood Ratio (LLR), Log-Area Ratio (LAR), Weighted Spectral Slope Distance (WSSD), Perceptual Evaluation of Speech Quality (PESQ) (ITU-T P.862), Sprachqualitätsmaß „ q_C “, Audioqualitätsschätzmethode PEMO-Q. Die Maße ISD, LLR und LAR basieren auf dem Vergleich von Sprachmodellparametern, die mit Hilfe der Methode der linearen Prädiktion gewonnen werden [3]. Das WSSD-Maß verwendet eine gehörorientierte Filterbank zur Spektralanalyse und vergleicht die Steigungsverläufe der Spektren von Test- und Referenzsignal [3]. PESQ [4], q_C [5] und PEMO-Q [6] verwenden Modelle der auditorischen Signalverarbeitung, um Test- und Referenzsignale auf ihre „interne Repräsentationen“ abzubilden und in dieser perzeptiven Ebene miteinander zu vergleichen. Alle aufgeführten modellbasierten Methoden benötigen zur Bewertung eines Testsignals immer auch ein Referenzsignal, das die optimale Qualität definiert.

Anwendung der Maße

Der hier vorliegenden Aufgabenstellung an die Qualitätsmaße und die Art der subjektiven Skalierung wurde bei der Qualitätsvorhersage Rechnung getragen, indem die Maße anders als üblich eingesetzt werden, und zwar so wie in Abb. 1 skizziert. Dadurch wird die durch den Beamformer bewirkte Qualitätsverbesserung gegenüber dem Eingangs-

signal abgeschätzt. Dies entspricht auch der subjektiven Messgröße.

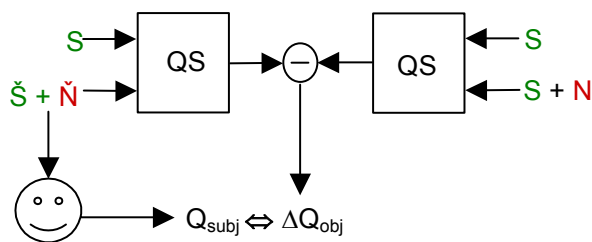


Abb. 1: Anwendung objektiver Qualitätsschätzer (QS) zur Vorhersage subjektiver Qualitätsbewertungen Q_{subj} . Zunächst werden Qualitätsschätzwerte für das verarbeitete Sprach-Störgeräuschgemisch ($\check{S}+\check{N}$) und das unverarbeitete Sprach-Störgeräuschgemisch ($S+N$) berechnet. Als Referenz dient jeweils das rauschfreie, originale Sprachsignal (S). Die Differenz der beiden Schätzwerte (ΔQ_{obj}) dient als Prädiktor der empfundenen Qualitätsverbesserung durch die jeweiligen Beamformer.

Ergebnisse

Tab. 1 zeigt Korrelationen zwischen gemessenen und geschätzten Qualitätswerten. Demnach korrelieren die SNR-Maße nur mäßig mit den subjektiven Gesamtqualitätsurteilen, selbst wenn stark artefaktbehaftete Signale, die durch zwei Beamformer mit fehlerhaften automatischen Sprecherlokalisierungen verursacht wurden, unberücksichtigt bleiben (s. Spalte 4 und 5). Dies ist insofern etwas überraschend, als dass man bei der hohen Sprachqualität der durch die Beamformer verarbeiteten Signale annehmen sollte, dass die Gesamtqualität vor allem durch den Störgeräuschhintergrund bestimmt sein dürfte, der sehr viel stärker zwischen den Testsignalen variiert als das Sprachsignal. Im Gegensatz zu den SNR-Maßen zeigen die meisten der stärker wahrnehmungsorientierten Maße z.T. deutlich bessere Übereinstimmungen mit den subjektiven Daten. Die höchste Korrelation erzielt das PEMO-Q-Maß PSM („Perceptual Similarity Measure“), aber auch die Maße q_c und WSSD zeigen recht gute Vorhersageleistungen.

Abb. 2 zeigt einen Scatterplot subjektiver vs. objektiver Qualitätswerte (ΔPSM). In rot dargestellt sind Qualitätsbewertungen der o.g. fehlerhaften Beamformer. Sie erhielten negative subjektive Skalenwerte, d.h. ihre Gesamtqualität wurde schlechter als die der unverarbeiteten Signale bewertet. Es lässt sich erkennen, dass das Ausmaß der empfundenen Störung durch die Verarbeitungsartefakte durch das objektive Qualitätsmaß teilweise noch unterschätzt wird.

Fazit

Insgesamt lässt sich festhalten, dass die Beurteilung der Gesamtqualität der hier untersuchten mehrkanaligen Störgeräuschreduktionsverfahren in Pkw durch gehörorientierte objektive Qualitätsmaße möglich ist und z.T. sehr gute, deutlich bessere Ergebnisse liefert als die Betrachtung der SNR-Verbesserung, selbst bei sehr artefaktarmer Verarbeitung und vermeintlich SNR-bestimmter Gesamtqualität.

Maß	alle BMF		ohne fehlerhafte BMF	
	lin. Kor.	Rangkor.	lin. Kor.	Rangkor.
SNR	0,71	0,68	0,65	0,49
segSNR	0,60	0,61	0,72	0,63
ISD	-0,62	-0,77	0,64	-0,66
LLR	-0,83	-0,91	-0,87	-0,87
LAR	-0,78	-0,80	-0,75	-0,60
WSSD	-0,87	-0,89	-0,88	-0,86
PESQ	0,83	0,79	0,70	0,63
q_c	0,87	0,90	0,93	0,90
PEMO-Q	0,90	0,94	0,95	0,93

Tab. 1: Korrelationen zwischen subjektiven Qualitätsmesswerten (BTL-Skalenwerte) und vorhergesagten Werten durch objektive Qualitätsmaße. „Lin. Kor.“: Linearer Korrelationskoeffizient; „Rangkor.“: Rangkorrelationskoeffizient nach Spearman; „BMF“: Beamformer

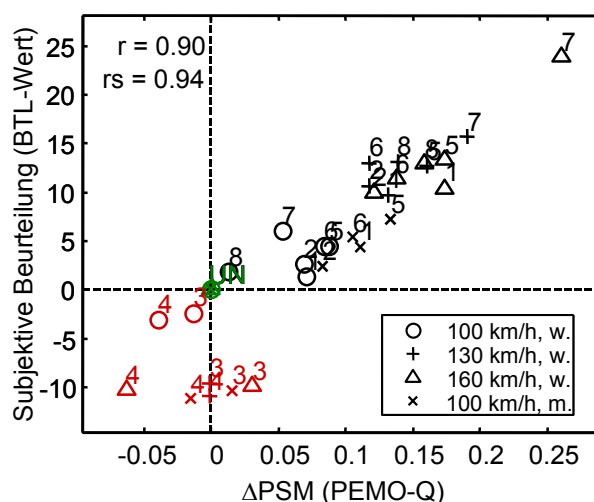


Abb. 2: Qualitätsvorhersage auf Basis des PEMO-Q-Qualitätsmaßes PSM. Symbole: Testkonditionen. Ziffern: Beamformer. „UN“: unverarbeitete Signale.

Literatur

[1] Tontch, K. „Instrumentelle und auditive Beurteilung der Qualität des Sprachsignals nach räumlicher Filterung des Sprechschalls in Personenkraftwagen“, Shaker Verlag, ISBN 3-8322-0411-3, 2002.

[2] Luce, R. D. "The choice axiom after twenty years." J.Math. Psych. **15** (1977), 215-233.

[3] Hansen, J. and B. Pellom. „An effective quality evaluation protocol for speech enhancement algorithms.“ Proceedings ICSLP 1998, Sydney, Australia.

[4] ITU-T. „Perceptual evaluation of speech quality (PESQ), an objective method for end-to-end speech quality assessment of narrow-band telephone networks and speech codecs.“ Recommendation P.862. (2001), Geneva, International Telecommunications Union.

[5] Hansen, M. and B. Kollmeier. "Objective modelling of speech quality with a psychoacoustically validated auditory model." J. Audio Eng. Soc. **48** (2000): 395-409.

[6] Huber, R. and B. Kollmeier. „Temporal aspects in the prediction of perceived audio quality differences.“ Joint Congress CFA/DAGA 2004, Strasbourg, France, Deutsche Gesellschaft für Akustik e.V.