# Mehrkanalige Sprachsignalerfassung im Kfz

**Kilian Tontch[1], Gordon Seitz[2], Dr. Klaus Schaaf[2]**

Volkswagen AG, [1] Elektronik-Forschung, [2] Fahrzeug-Forschung
Brieffach 1776, D-38436 Wolfsburg, Tel.:+49-5361-936233, Fax.:+49-5361-936709
e-mail: Kilian.Tontch@Volkswagen.de

## 1. Motivation

Die in Zukunft stark zunehmende Verwendung gesetzlich vorgeschriebener Telefonfreisprecheinrichtungen, sowie der komfort- und sicherheitsbedingte Einsatz automatischer Spracherkennungssysteme in modernen Kraftfahrzeugen zur einfacheren Bedienung multifunktionaler Fahrerinformationsysteme während der Fahrt, machen den Einbau von mindestens einem Mikrofon je Fahrzeug erforderlich. In der stark störgeräuscherfüllten Fahrzeugumgebung mit ihren vielfältigen unstetigen Randparametern, ist der Einsatz von Mikrofonarrays besonders erfolgversprechend, denn bei mehrkanaligen Mikrofonanlagen kann eine stark ausgeprägte, sprecherselektive und variable Richtcharakteristik eingestellt werden [1-3]. Die Eignung von Mikrofonanordnungen für den Einsatz im Kfz und die Leistungsfähigkeit verschiedener Beamformingverfahren soll untersucht werden. Hierbei wird der Artikulationsindex (AI) zur vergleichenden Beurteilung der Verfahren verwendet.

## 2. Sprachdatensammlung im Kfz

Im Rahmen von Untersuchungen zu mehrkanaliger Sprachsignalgewinnung im Kfz wurden mehrere verschiedene vielkanalige Sprachdatenbasen mit bis zu zweiundzwanzig aufgezeichneten Kanälen erstellt. Nach vorangegangenen grundlegenden Untersuchungen zu der Fahrzeuginnenraumakustik, wurden fahrzeugtypische Sprech- und Störschallsituationen gezielt und unter kontrollierten Bedingungen in fahrenden Fahrzeugen inszeniert und aufgezeichnet. Die Auswahl der Szenarien erfolgte unter den Gesichtspunkten der Realitätsnähe und im Hinblick auf akustisch anspruchsvolle Störsituationen. Die Vielfalt der fahrzeugtypischen Störgeräusche wird durch sechs Geschwindigkeitsbereiche und die Verwendung von vier Fahrzeugtypen in der Datenbasis abgebildet. Wechselnde Abstände zwischen der Signalquelle und den Mikrofonen, sowie drei Störsprecherszenarien runden die Datenbasis ab. Eine der Mikrofonanordnungen geht aus der Kombination von zwei linearen Arrayanordnungen hervor, eine andere aus einer Rechteckanordnung [4]. Für die Meßdatenerfassung wurden Präzisionsmikrofone, ein Kunstkopf und geeignete Multikanal-Meßsysteme eingesetzt. Als Testsatz dienen die Worte „Morgenstund hat Gold im Mund.“ Sie werden von einer Frauen- und von einer Männerstimme nacheinander artikuliert. Dazwischen ist eine ca. 2 Sekunden lange Pause, sodaß auch reines Fahrgeräusch aufgezeichnet wird.

## 3. Beamformingverfahren

Man unterscheidet zwischen mehrkanaligen Störschallkompensations- und Störschallreduktionsverfahren. Hier werden ausschließlich Störschallreduktionsverfahren betrachtet, die ohne Referenzmikrofon für den Störschall auskommen. Der Ausgleich der Laufzeitunterschiede des Sprechschalls zu den im allgemeinen unterschiedlich weit von dem Sprechermund entfernten Einzelmikrofonen eines Arrays ist ein wesentlicher Bestandteil jedes Beamformers. Für die vergleichenden Untersuchungen wurden der grundlegende Delay-and-Sum Algorithmus (DSB), der Filter-and-Sum Algorithmus und als adaptive Verfahren der Frost und der Griffiths-Jim-Beamformer implementiert (unter MATLAB). Die Sprachdaten aus den Fahrzeugen können so im Labor mit den verschiedenen Beamformingverfahren mehrfach bearbeitet werden. Durch die Kenntnis der exakten Abstandsmaße und Winkel zum Zeitpunkt der Aufnahmen, kann die Laufzeitkompensation sowohl manuell vorgegeben werden, als auch adaptiv über ein fahrzeuggeräusch- und störsprecherrobustes Ortungsverfahren bestimmt werden (Super-Resolution) [4]. Fehler bei der Ausrichtung auf die Nutzsignalquelle, wie sie z.B. bei einer bewegten Schallquelle auftreten können, führen zu Verzerrungen des Beamformerausgangssignals.

## 4. Beurteilungsverfahren

Das Verfahren der auditiven Beurteilung der Signale durch Testpersonen ist sehr aufwendig und zeitintensiv, so daß sich maschinelle Verfahren anbieten. Die Bestimmung des Signal-zu-Rauschabstands erlaubt keine psychoakustisch motivierten Interpretationen. Mit der aus nur sehr wenigen unterschiedlichen Worten bestehenden Datenbasis ist auch kein Spracherkennungssystem zum späteren Vergleich von Erkennerraten trainierbar. Die Verwendung des Artikulationsindex bietet sich als ein Maß für die Sprachverständlichkeit an. Der AI wird durch Terzanalysen des zu beurteilenden Geräusches und eines mittleren genormten Sprachpegels, sowie über Differenzbildung der Terzanteile, und durch die frequenzselektive Gewichtung (unter psychoakustischen Gesichtspunkten) der einzelnen Bänder und durch abschließende Summation aller so bestimmten frequenzbandabhängigen Anteile berechnet [5]. Der AI stellt dann ein Maß für die sprachverdeckende Eigenschaft des Geräusches dar. Wenn ein Geräusch gleichzeitig auch Sprache enthält, dann weist der AI an diesen Stellen geringere Werte auf, denn Sprache verdeckt

naturgemäß auch andere Sprachsignale. Genau dieser Effekt wird hier zur vergleichenden Beurteilung von Beamformern ausgenutzt. Bei Differenzbildung der AI-Werte zweier verschiedener Beamformerausgangssignale desselben Sprachsegments, entspricht die Differenz dem Verständlichkeitsunterschied der Beamformer und soll als Vergleichskriterium dienen.

Als Beispiel sei ein Sprachsignal aus einem Mittelklassefahrzeug bei v=100 km/h betrachtet. Die neun Mikrofone sind linear vor dem Sprecher angeordnet. Als Referenzsignal wurde das über den Kunstkopf abgestrahlte Signal gewählt. Verglichen werden die Ausgangssignale eines Delay-and-Sum Beamformers mit externer Laufzeitkompensation, und ein mit automatischer Ortung ausgerichteter Beamformer. In **Abbildung 2** sind die Originalsignale dargestellt. In **Abbildung 1** ist der Unterschied der Beamformingvarianten direkt als auf das Referenzsignal bezogene Differenz aufgetragen. Während in der reinen Fahrgeräuschphase kaum ein Unterschied erkennbar ist (t=0-0.7 s) (Abb. 2), weist der Delay-and-Sum Beamformer in den Sprachabschnitten höhere Werte für den AI auf als der Ortungs-Beamformer (Abb. 2). Das stimmt mit dem besseren Höreindruck für den DSB überein, und findet sich in Abbildung 1 als höherer Wert. Die negativen Werte rühren von dem idealen Referenzsignal her. Relevant ist der Unterschied zwischen den Kurven.

## 5. Zusammenfassung und Ausblick

Anhand von mehrkanaligen Sprachdatenbasen werden verschiedene Mikrofonarrays und Beamformingverfahren mit Hilfe des Artikulationsindex verglichen. Dabei wird der AI zur Beurteilung des Unterschieds verschiedener Beamformingmethoden herangezogen.

Der nächste Schritt wird die differenziertere Auswertung unter Berücksichtigung aller verfügbaren Parameter der Datenbasis sein. Desweiteren ist geplant, die Datenbasis um eine nichtlineare halbkreisförmige Mikrofonanordnung mit drei unterschiedlichen Radien zu erweitern. Das beschriebene Verfahren ist evtl. durch die Verwendung des Referenzsignals als mittlerer Sprachpegel bei der Bestimmung des AI ausbaufähig.

## Literatur


[1] Kroschel K., Lange K.: „Einsatz von Mikrofonarrays für Freisprecheinrichtungen im Kraftfahrzeug“, Kleinheubacher Berichte, Tagungsband, Okt. 1994

[2] Van Compernolle D. et al.: „Speech recognition in noisy environments with the aid of microphone arrays“, Speech Communication 9, Elsevier Science Publishers B.V. (North-Holland), S. 433-442, 1990

[3] Smolders J., Claes T., Sablon G., Van Compernolle D.: „On the importance of the microphone position for speech recognition in the car.“, Proc. ICCASP, Vol. 1, S. 429-432, 1994

[4] Tontch K., Wolf M., Seitz G., Schaaf K.: „Sprecherortung im Kraftfahrzeug mit Hilfe des Super-Resolution Ansatzes“, Tagungsband Sprachkommunikation, Ilmenau, Okt. 2000

[5] Accredited Standards Committee S3, Bioacoustics: „Methods for the calculation of the articulation index“, American National Standards Institute, ANSI S3.5, 1969


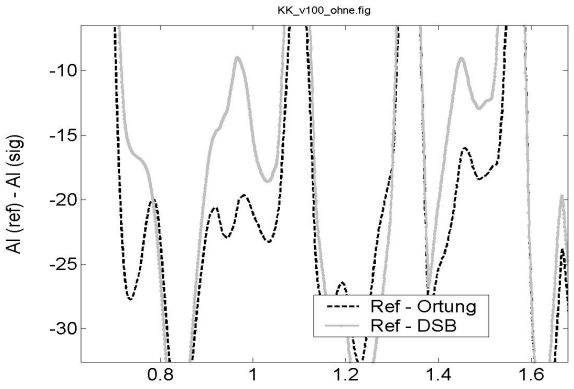


**Abbildung 1: Aufgetragen ist die Differenz des Artikulationsindex zwischen dem: (1) Referenzsignal und dem Ortungs-Beamformer (schwarz) und (2) Referenzsignal und Delay-and-Sum-Beamformer (grau). Die Unterschiede im AI der Beamformersignale werden ersichtlich.**

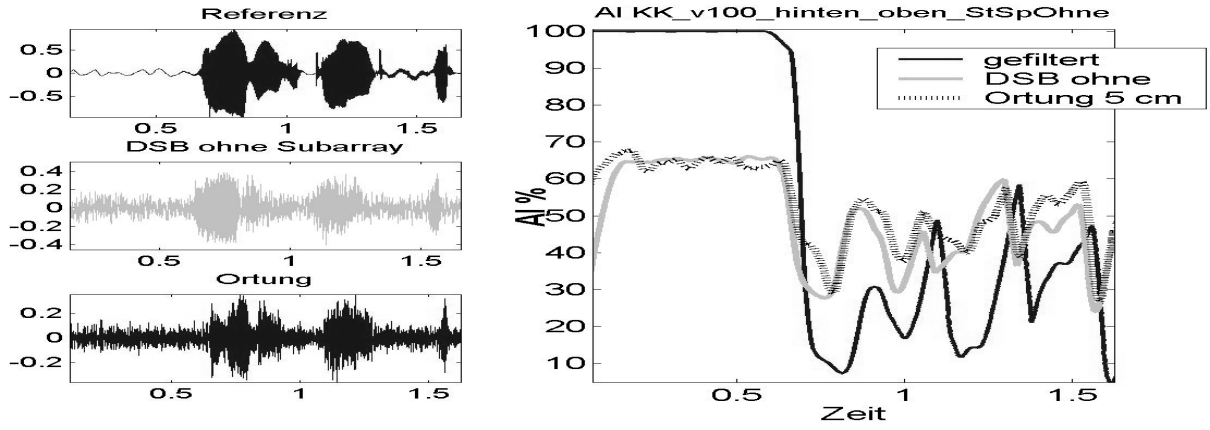


**Abbildung 2: Darstellung des Referenzsignals und der Ausgangssignale zweier Beamformer bei v=100 km/h im Zeitbereich (links), sowie des AI für diese Signale. (Referenz: schwarz, DSB-BMF: grau, Ortung: punktiert).**