

Untersuchung von Gradientenmikrofonen höherer Ordnung für automatische Spracherkennung in Kraftfahrzeugen

Roland Aubauer und Dieter Leckschat

Siemens AG, ICP CD TI 2, Frankenstr. 2, 46395 Bocholt, Email: roland.aubauer@bch.siemens.de

Zusammenfassung

Es wird ein Gradientenrichtmikrofon zweiter Ordnung vorgestellt, das im Hinblick auf Freisprechen und automatische Spracherkennung in Kraftfahrzeugen untersucht wurde. Die Entwicklung eines einfachen elektronischen Empfindlichkeitsabgleichs ermöglicht die Verwendung von preisgünstigen Mikrofonkapseln mit üblichen Toleranzen. Die Richtcharakteristik ist im Telefonfrequenzbereich nahezu frequenzunabhängig. Das theoretisch erzielbare Bündelungsmaß des Mikrofons ist mit 8,8 dB deutlich höher als das von konventionellen Richtmikrofonen, die in der Telekommunikation eingesetzt werden. Störgeräusche und unerwünschte Schallanteile, die aus Richtungen außerhalb der Mikrofonhaupttrichtung eintreffen, wie z. B. Nachhall und konkurrierende Sprecher werden signifikant stärker unterdrückt. Das Mikrofon wurde in einer Kraftfahrzeugumgebung anhand der Erkennungsrate eines sprecherabhängigen Einzelworterkenners mit Gradientenmikrofonen geringerer Ordnung verglichen. Bei einem optimalen Einbau konnte die Störgeräuschrobustheit des Erkenners um mehr als 4,7 dB gesteigert werden. Die Einflüsse der Raumakustik und der Mikrofonpositionierung auf die Erkennungsrate wurden gesondert untersucht. Es konnte gezeigt werden, dass der in einem Fahrzeug auftretende diffuse Nachhall praktisch keinen Einfluss hat, jedoch einzelne, starke Reflexionen die Spracherkennungsrate in einer Störgeräuschkulisse deutlich beeinträchtigen können. Es werden Schlüsse bezüglich einer optimalen Anordnung von Freisprechmikrofonen in Kraftfahrzeuge gezogen.

Einleitung

Freisprechtelefonie und sprachgesteuerte Systeme finden im Kraftfahrzeug zunehmend an Bedeutung. Diese Systeme dienen dabei nicht nur der Bequemlichkeit des Benutzers, sondern sind ein maßgeblicher Sicherheitsaspekt. Heutige Spracherkennungssysteme zeichnen sich allerdings noch durch eine mangelnde Robustheit gegenüber Störgeräuschen, Raumhall und anderen Störungen des Übertragungskanal aus. Gleichzeitig fordern mehrere gleichzeitig aktive Schallquellen die Spracherkennungssysteme heraus. Im Freisprecheinsatz, bei dem das Sprachaufnahmehinrichtung weit vom Mund des Sprechers entfernt ist, wirken sich Störeinflüsse besonders dramatisch aus. Die in der Praxis erzielten Spracherkennungsergebnisse sind darum oft erheblich schlechter als die in den Sprachlabors unter idealen Bedingungen evaluierten Ergebnisse.

In diesem Bericht wird ein Gradientenmikrofon zweiter Ordnung (Quadrupolmikrofon) bezüglich seiner Toleranzanforderungen untersucht und in einem Kraftfahrzeug mit Gradientenmikrofonen geringerer Ordnung anhand von Spracherkennungsergebnissen verglichen. Es soll gezeigt werden, dass die mit dem Quadrupolmikrofon theoretisch zu erwartende Störfreie auch in der Praxis erreicht werden kann.

Gradientenmikrofone erster und zweiter Ordnung

Mikrofone, die Schall nur mit einer Seite einer Membran aufnehmen, sind Drucksensoren. Deren Richtcharakteristik entspricht einer Kugel, solange die Mikrofone klein gegenüber der Schallwellenlänge sind. Richtwirkung kann erzeugt werden, wenn beide Seiten einer Membran dem Schall zugänglich gemacht werden, so dass das Mikrofon empfindlich gegenüber dem Druckgradienten wird. Beispiele dieser Wandler sind Dipol- und Nierenmikrofone (Gradientenmikrofone erster Ordnung).

Ein Gradientenmikrofon zweiter Ordnung mit verbesserter Richtcharakteristik und Störgeräuschunterdrückung wird durch die Kombination zweier Nierenmikrofone erreicht, deren Ausgangssignale voneinander subtrahiert werden (Sessler und West, 1975, Olson, 1979). Die Mikrofone werden hierzu längs ihrer Rotationsachsen in einem Abstand Δx (in unserem Fall gleich 3 cm) angeordnet. Die Ausgangssignale der Nierenmikrofone sind:

$$x(t, \varphi) = \Gamma_c(\varphi) \cdot \hat{p} e^{j2\pi f t}, \quad (1)$$

$$y(t, \varphi) = \Gamma_c(\varphi) \cdot d \cdot \hat{p} e^{j2\pi f (t + (\Delta x/c) \cdot \cos \varphi)}, \quad (2)$$

wobei \hat{p} die Amplitude, φ der Schalleinfallswinkel, $\Gamma_c(\varphi)$ die Nierenrichtcharakteristik, c die Schallgeschwindigkeit, f die Frequenz und d ein Empfindlichkeitsfaktor zwischen den Nierenmikrofonen sind. Das Differenzsignal $s(t, \varphi) = x(t, \varphi) - y(t, \varphi)$ ist für $f \ll c/\Delta x$:

$$s(t, \varphi) \approx \Gamma_c(\varphi) \cdot ((1-d) - j \cdot (d \cdot 2\pi f \cdot (\Delta x/c) \cdot \cos \varphi)) \cdot \hat{p} e^{j2\pi f t}. \quad (3)$$

Bei identischer Empfindlichkeit der Nierenmikrofone ($d=1$) ergibt sich die ideale Richtcharakteristik des Quadrupolmikrofons aus der Multiplikation der Nieren- mit der Dipolcharakteristik:

$$\Gamma_q(\varphi) = \Gamma_c(\varphi) \cdot \cos \varphi = 0.5 \cdot (1 + \cos \varphi) \cdot \cos \varphi \quad (4)$$

Empfindlichkeitsunterschiede zwischen den Nierenmikrofonen senken in der Praxis die Richtwirkung des Quadrupolmikrofons in Abhängigkeit der Signalfrequenz. Für $d \gg 1$ oder $d \ll 1$, wiegt der $(d-1)$ Term in Gl. (3) über den Imaginärterm $(d \cdot 2\pi f \cdot (\Delta x/c) \cdot \cos \varphi)$, so dass sich die Richtcharakteristik des Quadrupol- der eines Nierenmikrofons annähert. Die typische Quadrupol-Richtcharakteristik bildet sich bei einer angenommenen Empfindlichkeitsdifferenz der Mikrofonkapseln von 3 dB erst ab zirka 2000 Hz aus. Unterhalb von 500 Hz wird gegenüber einem Nierenmikrofon praktisch kein Gewinn bezüglich der Richtcharakteristik und der damit verbundenen Störgeräuschunterdrückung erzielt. Dieses Ergebnis ist bemerkenswert, da für einen Einsatz im Kraftfahrzeug eine Störgeräuschunterdrückung gerade im tieffrequenten Bereich gewünscht wird. Ein präziser Empfindlichkeitsabgleich der Nierenmikrofone ist so absolut notwendig.

Diese Problematik wurde durch einen automatischen Empfindlichkeitsabgleich gelöst. Aus der Differenz der Hüllkurven der Nierenmikrofonensignale kann ein Steuerungssignal zur Ausregelung der Empfindlichkeitsunterschiede generiert werden. Der obere Teil des in Abb. 1 gezeigten Blockschaltbild zeigt (in Grau) die generelle Funktionsweise des realisierten Quadrupolmikrofons, wohingegen der untere Teil (in Weiss) den automatischen Empfindlichkeitsabgleich darstellt. Der 500 Hz Tiefpass dient hier zum Ausgleich des frequenzproportionalen Amplitudenfrequenzgangs (Gl. 3, $d=1$).

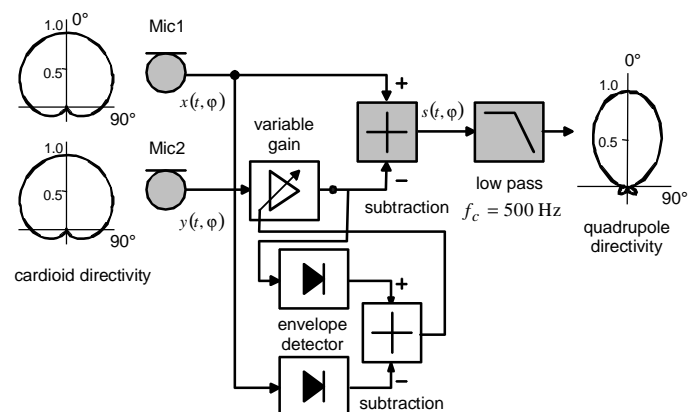


Abb. 1: Blockschaltbild des Quadrupolmikrofons

Das realisierte Quadrupolmikrofon mit automatischem Empfindlichkeitsabgleich weist eine deutlich verbesserte Richtcharakteristik im Frequenzbereich von 300 – 3400 Hz auf. Das die Unterdrückung von diffusem Störgeräusch gegenüber einem aus der Mikrofonhauptachse einfallenden Nutzsinal charakterisierende Bündelungsmaß ist mit $di \geq 8$ dB mindestens 2 dB höher als das eines idealen Hypernierenmikrofons ($di = 6$ dB).

Mikrofonuntersuchung im Kraftfahrzeug

Das Quadrupolmikrofon wurde in einem Kraftfahrzeug (Mercedes A-Klasse) an verschiedenen Montagepunkten (am Armaturenbrett, an der A-Säule oben, am Dachhimmel und an der Innenraumbeleuchtung) im Vergleich zu einem Nieren-, einem Hypernieren- und einem Kugelmikrofon untersucht. Hierzu wurden die kombinierten Mikrofon – Raumübertragungsfunktionen mittels Kunstkopf- und Maximalfolgenmesstechnik vom Mund des Kunstkopfes zum Ausgang des Mikrofons gemessen. Abbildung 2 zeigt typische Übertragungsfunktionen für zwei Mikrofone und zwei Positionen. Die zum Teil deutlichen Einkerbungen im Frequenzgang resultieren von Interferenzen des Direktsignals mit Mehrwegeausbreitungen, z. B. über die Seitenfenster des Fahrzeugs. Das Quadrupolmikrofon unterdrückt den Schalleinfall außerhalb der Haupteinfallrichtung effektiv, so dass Einkerbungen der Übertragungsfunktion deutlich verringert werden. In der Position 'Dachhimmel' (Mikrofon – Mundabstand 0,19 m) wird der Frequenzgang der Gradientenmikrofone durch den Nachbesprechungseffekt im unteren Frequenzbereich deutlich angehoben, so dass der Signal-/Rauschabstand zusätzlich verbessert wird.

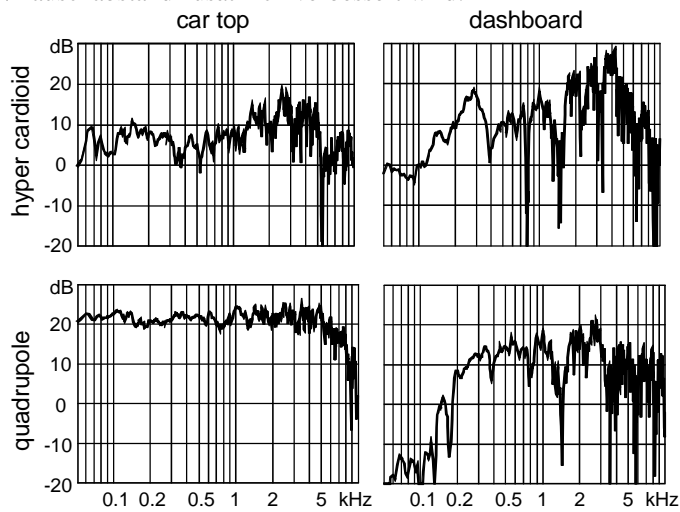


Abb. 2: Kombinierte Mikrofon-Raumübertragungsfunktionen

Das Fahrgeräusch wurde mit allen vier Testmikrofonen an allen Positionen (4x4) bei drei verschiedenen Geschwindigkeiten gemessen (80, 100, 130 km/h). Tieffrequente Anteile des Wind- und des Reifenabrollgeräuschs dominieren das Fahrgeräuschspektrum, welches bis etwa 1 kHz mit ungefähr 6 dB und darüber mit ungefähr 12 dB pro Oktave abfällt. Die Mikrofonposition 'Innenraumbeleuchtung' zeichnet sich durch das geringste aufgenommene Geräusch aus, wohingegen die 'A-Säule' relativ laut ist. Das Quadrupolmikrofon nimmt gegenüber dem Kugelmikrofon fast frequenzunabhängig 10 dB und gegenüber den Nierenmikrofonen zwischen 3 und 5 dB weniger Rauschen auf, bezogen auf die gleiche Mikrofonempfindlichkeit in Haupttrichtung (Aubauer et. al. 1999).

Die Eignung der Mikrofone für den Freisprecheinsatz im KFZ wurde mit Hilfe eines sprecherabhängigen Einzelwort-Spracherkenners untersucht. Die Erkennung basiert auf einem Vergleich der eingegebenen Sprache mit den während einer Trainingsphase aufgenommenen Sprachaufnahmen anhand Cepstralkoeffizienten und mit Hilfe der Methode der dynamischen Programmierung (Dynamic Time Warping). Beginn und Ende eines Wortes werden durch automatische Sprachaktivitätserkennung bestimmt. Das Spracherkennungssystem wurde mit je zwei Aussprachen eines 20 Worte Vokabulars von je drei weiblichen und männlichen Sprechern trainiert. Die Spracherkennung wurde mit einer dritten Aussprache getestet.

Die einer Sprachdatenbank entnommenen Aufnahmen wurden mit den Mikrofon-Raumimpulsantworten gefaltet, um Sprachaufnahmen im Fahrzeuginnenraum zu simulieren. Für die Er-

kennung wurde zusätzlich das aufgenommene Fahrgeräusch unterschiedlicher Geschwindigkeiten zugemischt. Die gewählte Vorgehensweise bietet gegenüber echten Sprachaufnahmen mit verschiedenen Sprechern im KFZ den Vorteil absolut konstanter Testbedingungen, d. h. die Mikrofone können mit exakt den gleichen Sprachäußerungen und den gleichen Fahrgeräuschaufnahmen getestet werden.

Untersuchungsergebnisse

Abbildung 3 zeigt die Erkennungsergebnisse in Abhängigkeit des Fahrgeräuschpegels im Fahrzeuginnenraum. Die höchsten Erkennungsraten wurden unabhängig von der Montageposition mit dem Quadrupolmikrofon erzielt. Im Vergleich zum Kugelmikrofon konnte die Störgeräuschrobustheit des Erkenners zwischen 3,5 und 8,2 dB und im Vergleich zu den Nierenmikrofonen zwischen 0,9 und 4,8 dB verbessert werden. Die höchste Verbesserung wurde dabei in der insgesamt besten Montageposition am Dachhimmel erzielt. Die Kombination von Richtcharakteristik und Nahbesprechung des Quadrupolmikrofons wirkt sich in dieser Position besonders vorteilhaft aus.

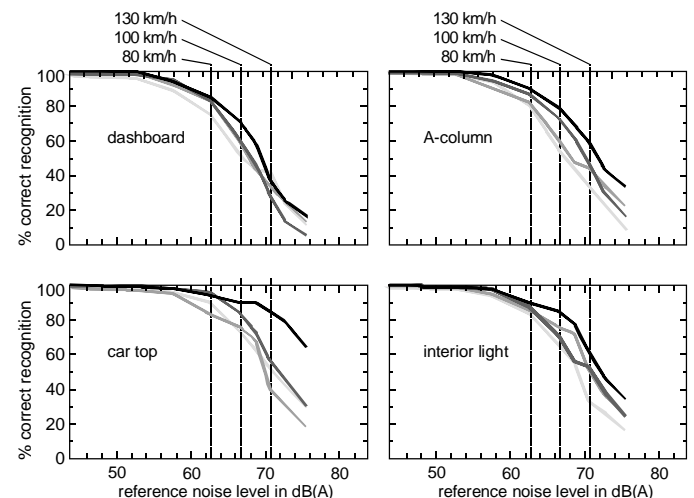


Abb. 3: Worterkennungsraten der simulierten Sprachaufnahmen mit den getesteten Mikrofonen (Quadrupol schwarz, Niere dunkelgrau, Hypernieren grau, Kugel hellgrau) an den verschiedenen Positionen.

Der Abstand des Mikrofons vom Mund des Sprechers (MMA) ist der wichtigste Parameter für die Mikrofonpositionierung im KFZ, da dieser direkt in den Störabstand des Sprachsignals eingeht. Die höchsten Erkennungsraten werden am Dachhimmel (MMA 0,19m) und die geringsten auf dem Armaturenbrett (MMA 0,7 m), unabhängig vom Mikrofontyp, erzielt. Mittlere Erkennungsraten werden mit den Mikrofonen an der Innenraumbeleuchtung (MMA 0,28m) und an der A-Säule (MMA 0,36m) erzielt, wobei die Innenraumbeleuchtung auch aufgrund geringerer Signalinterferenzen mit Mehrwegeausbreitungen tendenziell besser abschneidet. Der diffuse Nachhall im KFZ spielt bei der Spracherkennung keine Rolle.

Ausreichend hohe Erkennungsraten größer 85% werden bei 130 km/h nur mit dem Quadrupolmikrofon am Dachhimmel und bei Geschwindigkeiten bis 100 km/h an der Innenraumbeleuchtung erreicht.

Literatur

- H. F. Olson (1979), "Directional microphones", An anthology of articles on microphones from the pages of the Journal of the Audio. Eng. Soc.. Vol. 1-Vol. 27, (1953-1979), pp. 190-194.
- G. M. Sessler and J. E. West (1975), "Second-order gradient unidirectional microphone utilizing an electret transducer", Journal of the Acoust. Soc. Am. Vol. 58(1), pp. 273-278.
- R. Aubauer, R. Kern, D. Leckschat (1999), "Optimized second order gradient microphone for hands-free speech recordings in car", Proceedings on Robust Methods for Speech Recognition in Adverse Conditions, Tampere, Finland, pp.191-194