

Effiziente Verfahren zur Störgeräuschunterdrückung in Magnetresonanztomographen bei Sprachaufnahmen

M. Kob¹, G. Behler², R. Schnitker³

¹ Lehr und Forschungsgebiet für Phoniatrie und Pädaudiologie, RWTH Aachen – Universitätsklinikum, mkob@ukaachen.de

² Institut für Technische Akustik, RWTH Aachen, gkb@akustik.rwth-aachen.de

³ Service-Einheit funktionelle Bildgebung im IZKF Biomat., RWTH Aachen – Universitätsklinikum, schnitker@ukaachen.de

Einleitung

Die simultane Erfassung von Sprache und der Physiologie der Stimm- und Sprechorgane oder der Hirnaktivierung im Magnetresonanztomographen (MRT) ist für die Beurteilung der zeitlichen und inhaltlichen Zusammenhänge zwischen dem Stimmsignal und den Bilddaten von großer Bedeutung. Ein Problem bei der Aufnahme sind die während der Messung auftretenden, sehr lauten Störgeräusche des MRT als Folge der Gradientenschaltungen, die dem Sprachsignal bei einer Aufnahme im Gerät mit vergleichbarem oder sogar höherem Schallpegel überlagert sind. Besonders die schnelle Echo-Planar-Imaging (EPI) Technik, die fast ausschließlich als Basis für die funktionelle Bildgebung verwendet wird, macht es aufgrund des sehr lauten Störschallpegels unmöglich, verständliche sprachliche Reaktionen von Probanden und Patienten aufzuzeichnen. Der Signal-zu-Rauschabstand beträgt bei typischen Aufnahmen am Mund der Probanden/Patienten nur wenige dB. Oftmals ist es aber für die Auswertung der funktionelle MRT-Daten notwendig, das genaue zeitliche Auftreten von bestimmten sprachlichen Ereignissen (Versprecher, Pausen, Sprachpathologien) zu erfassen. Für eine spätere Analyse des Signals ist zwar keine Bereinigung des Sprachsignals in Echtzeit notwendig, doch wäre für einen interaktiven Test mit simultaner Beurteilung des Sprachsignals und Aufzeichnung der Hirnaktivitäten eine Geräuschsignalkompensation während der Versuchsdurchführung sehr wünschenswert.

Eine generelle Besonderheit des MRT-Störgeräuschs ist seine mit hoher Reproduzierbarkeit auftretende Periodizität über die Dauer der Messung hinweg. Da das Sprachsignal diese Eigenschaft nicht aufweist, kann dieser Unterschied für eine effiziente Reduktion des Störanteils genutzt werden. Erschwerend ist jedoch die ausgesprochen inhomogene Schallfeldverteilung im MRT-Gerät, die Schwankungen des Schalldruckpegels von über 10 dB je nach Ort des Mikrofons bewirkt.

In diesem Beitrag werden verschiedene Verfahren in einem vorläufigen subjektiven Test verglichen, die den Störanteil der Sprachaufnahme bei MRT-Untersuchungen reduzieren sollen.

Methoden

Zur Reduzierung des Scannergeräuschs lassen sich Spektral- und Zeitbereitsverfahren unterscheiden.

Spektrale Verfahren

Ein für die Geräuschreduktion wichtiger Aspekt ist die Spektralverteilung des Störsignals, die bei der EPI-Technik der Charakteristik eines Vokalspektrum sehr ähnlich ist. In Abbildung 1 ist ein Vokalspektrum dem MRT-Geräuschspektrum an jeweils gleicher Aufnahmeposition im Mundbereich eines Probanden dargestellt. Ein Ansatz zur Reduktion des Geräusch im

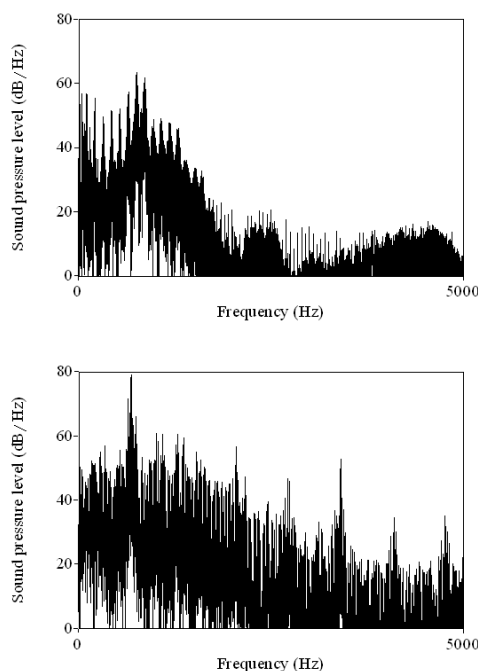


Abbildung 1: Oben: Spektrum des Vokals /a:/; Unten: MRT-Frequenzspektrum bei EPI-Protokoll (TR=3000ms, TE=30ms, 3 Tesla): harmonische Obertonstruktur, nicht harmonische Komponenten sowie breitbandige Rauschanteile sind akustische Korrelate der Phasenkodier-, Frequenz- und Schichtselektionsgradienten.

Frequenzbereich ist die Entfernung aller zum Scanner-signal gehörigen harmonischen Komponenten. Ein Algorithmus, der auch zeitvariante Grundfrequenzverläufe berücksichtigt ist der PSHF-Ansatz (pitch scaled harmonic filter) [1]. Bei dem Verfahren wird von überlappenden Zeitabschnitten des Störgeräuschs nach einer Grundfrequenzbestimmung eine Entfernung einer vorgegebenen Anzahl der Harmonischen durchgeführt. Das Verfahren arbeitet für den harmonischen Anteil des Störgeräuschs zuverlässig, doch bleiben die breitbandigen und unharmonischen Anteile unberücksichtigt.

Zeitbereichs-Verfahren

Das konzeptionell einfachste Verfahren im Zeitbereich ist sicherlich die Subtraktion zweier Mikrofonaufnahmen, bei der die eine nur das Störsignal und die andere das aus Sprach- und Störanteil bestehende Signal enthält. Hierbei treten jedoch zwei Probleme auf. Zum einen muß gewährleistet sein, dass das Sprachsignal nicht auf den Störkanal überspricht, die beiden Signale jedoch hinsichtlich des Störanteils hochgradig korreliert sind. Dies ist bei zweikanaliger Aufnahme nur sehr schwer zu realisieren, da der für schalldämmende Maßnahmen zur Verfügung stehende Platz in den heute verwendeten Mehrkanal-Kopfspulen – vor allem in Kombination mit dem notwendigen Stimulationsequipment (Köpfungör, Videobrille) – nur sehr gering ist. Auch kann nicht davon ausgegangen werden, dass das Störgeräusch ausschließlich von außen an das Mikrofon dringt. Vergleichsmessungen an verschiedenen Positionen um den Kopf des Probanden zeigen, dass das Störgeräusch zum Isozentrum des Magneten hin sogar zunimmt.

Für die direkte Subtraktion bietet sich neben der zweikanaligen Aufnahme auch die Subtraktion eines zuvor ohne Sprachsignal aufgenommenen Referenzsignals an, da die Scannergeräusche sehr genau reproduziert werden (Bild 2). Wie die Ergebnisse zeigen, bedeutet dies jedoch

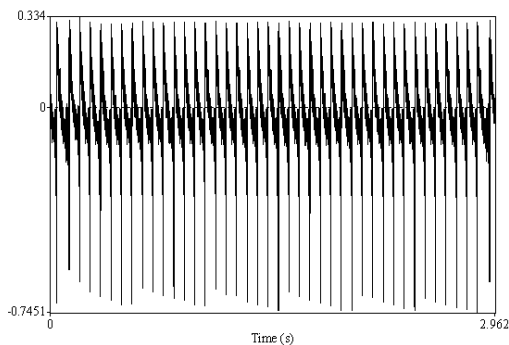


Abbildung 2: Zeitreihe eines EPI-Repetitionszyklus von ca. 3 s (TR) mit einer Echozeit von 30 ms. Die einzelnen Amplitudenspitzen repräsentieren die gemessenen 40 Schichten.

nicht, dass die beiden Mikrofonsignale auch hochgradig korreliert sind.

Als drittes Verfahren kann ein adaptives Filterverfahren eingesetzt werden, bei dem die Koeffizienten eines digitalen Filters abhängig von Sprach- und Störschallsignal optimiert werden. Besonders geeignet scheint hier das modellbasierte NLMS (normalized least-mean-squares)-Verfahren zu sein (Bresch et al. [2]). Vorteil dieses Ansatzes ist die Anpassungsfähigkeit des Verfahrens an die Charakteristik des Störsignals und die sehr hohe Störgeräuschbefreiung. Jedoch muss auch dieses Verfahren eine gute Vorgabe des Störsignals bekommen um effektiv zu sein, d.h. eine sprachunabhängige Mikrofonaufnahme des Störsignals verbessert die Adaptionsfähigkeit des Filters. Für die Anwendung bei der Lungengeräuschanalyse ist eine besondere Anordnung der Mikrofone recht wirksam [3], die jedoch für Sprachschallanalysen nicht anwendbar ist.

Ergebnisse

Das PSHF-Verfahren lieferte eine subjektiv verbesserte Sprachverständlichkeit, doch verblieb ein stark dominierendes Störsignal, das eine weitergehende Analyse der Sprachsignale unmöglich machte.

Die Subtraktion eines zuvor aufgenommenen Referenz-Störsignals arbeitete für kürzere Signalabschnitte sehr effektiv, doch wechselten sich gut verständliche Passagen nach jeweils wenigen Sekunden mit stark gestörte Abschnitten ab, in denen das Störsignal sogar stärker als im Fall ohne Subtraktion war. Dies könnte auf eine Laufzeitänderung des Signals durch Bewegung oder Temperaturänderung beim Sprechen zurückzuführen sein.

Die Anwendung des adaptiven Filters bei Sprachaufnahmen führt zu einer sehr effektiven Unterdrückung des Störgeräuschs in Sprechpausen und bei angehaltenen Sprachklängen. Im Anklang eines Sprachlautes hingegen überlagert sich ein Störanteil, der auf die Adaptionszeit des Filters zurückzuführen ist. Je nach Sprachsignal ruft diese Adaptionszeit einen Echo-ähnlichen Klang hervor, der die Stimme stark verfremdet und eine subjektive und objektive Analyse der Sprachsegmente stark erschwert.

Diskussion

Für funktionelle Bilgebungsstudien wird es notwendig sein, das Störsignal des MRT nicht nur post hoc sondern in Echtzeit aus dem Sprachsignal heraus zu filtern. Dies ermöglicht z.B. die Untersuchung von unterschiedlichen Aspekten sprachlicher Interaktion zwischen Gesprächspartnern, was besonders bei Patienten mit Sprach- und Sprechstörungen von Interesse ist.

Im direkten Vergleich der drei beschriebenen Verfahren hat sich als Echtzeitanwendung der Geräuschunterdrückung für fMRT-Versuche kein Verfahren als geeignet herausgestellt. Vielversprechend erscheinen analoge Verfahren [4] und das NMLS-Verfahren mit kürzeren Adaptionszeiten. Diese könnte durch Verbesserung des Modells und Hardwareimplementierung des Verfahrens erreicht werden.

Literatur

- [1] Jackson, P.J.B. & Shadle, C.H.: Frication noise modulated by voicing, as revealed by pitch-scaled decomposition, *J. Acoust. Soc. Am.*, 2000, **108**, 1421-1434.
- [2] Bresch, E., Nielsen, J., Nayak, K., & Narayanan, S.: Synchronized and noise-robust audio recordings during realtime magnetic resonance imaging scans, *J. Acoust. Soc. Am.*, 2006, **120**, 1791-1794.
- [3] Patel, S.B. et al.: An adaptive noise reduction stethoscope for auscultation in high noise environments, *J. Acoust. Soc. Am.*, 1989, **103**, 2483-2491.
- [4] Lukkari, T., Malinen, & J., Palo, P.: Recording speech during magnetic resonance imaging, *Proceedings of the 2nd International Workshop on Models and Analysis of Vocal Emissions for Biomedical Applications (MAVEBA)*, University of Florence, 2007 1-4.