

# Adaptive Verbesserung der Sprachverständlichkeit und Medienwiedergabe in Fahr- und Verkehrsgeräuschen

Jan RENNIES<sup>1</sup>, Henning SCHEPKER<sup>1,2</sup>, Aleksandra KUBIAK<sup>1</sup>, Simon DOČLO<sup>1,2</sup>

<sup>1</sup>Fraunhofer Institut für Digitale Medientechnologie IDMT, Projektgruppe Hör-, Sprach- und Audiotechnologie, Oldenburg, E-Mail: jan.rennies@idmt.fraunhofer.de

<sup>2</sup>Universität Oldenburg, Department Medizinische Physik und Akustik, Signalverarbeitung, Oldenburg

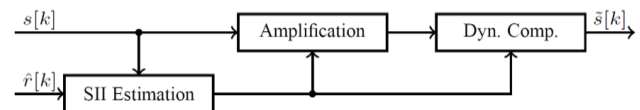
## Einleitung

Der empfundene Hörkomfort im Fahrzeuginnenraum wird neben der gewollten oder ungewollten Wahrnehmung von fahrzeugtypischen Funktions- und Betriebsgeräuschen auch dadurch bestimmt, wie angenehm dem Nutzer Sprache und Musik aus den diversen Medienkanälen dargeboten werden. Insbesondere in Anwesenheit von Fahrzeug- und Umgebungsgeräuschen können die Sprachverständlichkeit eingeschränkt und die bevorzugten Klangeinstellungen zur Musikwiedergabe unterschiedlich sein. Idealerweise sollte das Medienwiedergabesystem daher auf die sich über die Zeit ändernden Umgebungsgeräusche automatisch reagieren und zusätzlich die persönliche Nutzerpräferenz berücksichtigen. In dieser Studie werden zwei Verfahren vorgestellt, die die Klangwahrnehmung im Auto verbessern können: eine auf Grundlage von Sprachverständlichkeitsmodellen basierende störgeräuschadaptive Vorverarbeitung der Sprache, die ohne Erhöhung der Lautstärke eine bessere Verständlichkeit ermöglicht, und eine Nutzerschnittstelle zur Klangpersonalisierung, welche leichter und intuitiver zu bedienen ist als die üblichen Presets (Jazz, Klassik, etc.). Mithilfe dieser beiden Verfahren wurde die interindividuelle Variabilität und der Einfluss von Hintergrundgeräuschen auf individuelle Klangpräferenzen beim Musikhören im Auto untersucht.

## Adaptive Verständlichkeitsverbesserung

In [1] wurde der Algorithmus AdaptDRC (adaptive dynamic range compression) vorgestellt, welcher schematisch in Abb. 1 dargestellt ist. Ziel des Algorithmus ist eine Verbesserung des Sprachverstehens von Wiedergabesystemen, indem in Abhängigkeit von einem beim Nutzer vorliegenden Störgeräusch eine frequenzabhängige Verstärkungsregelung und Dynamikkompression verwendet wird. Als Steuergröße des Algorithmus dient eine vereinfachte Form des Sprachverständlichkeitsindex (SII). Das auszugebene Sprachsignal  $s[k]$  wird bei einem gegebenen Störgeräusch  $r[k]$  adaptiv gemäß Schätzungen des SII modifiziert, so dass in Situationen mit schlechter Verständlichkeit (z.B. laute Fahr- und Umgebungsgeräusche) das Sprachsignal am Ausgang zu einem rosa Spektrum geformt und zudem in seiner Dynamik komprimiert wird. Dies führt im Allgemeinen zu einer verstärkten Betonung hoher Frequenzanteile, was zu einer erhöhten Klarheit und einer erhöhten Sprachverständlichkeit führen soll. Bei hinreichend guter Verständlichkeit hingegen ist der Algorithmus transparent, d.h. es erfolgt keine Veränderung des

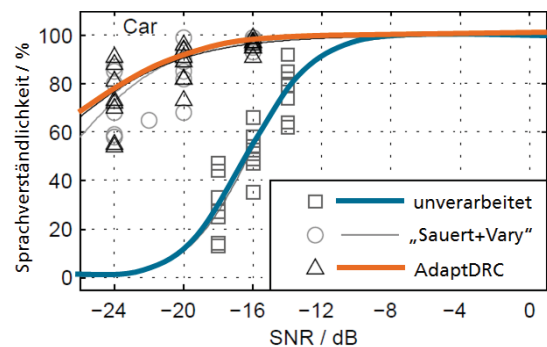
wiedergehenden Sprachsignals. Der Pegel am Eingang und Ausgang bleibt dabei erhalten, d.h. es erfolgt keine Erhöhung des Breitbandpegels.



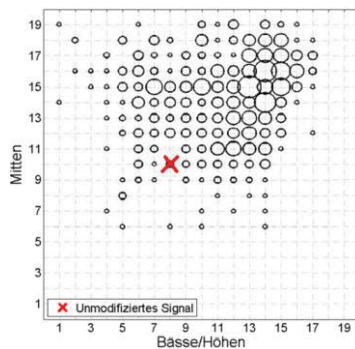
**Abbildung 1:** Schematische Darstellung des AdaptDRC-Verfahrens zur störgeräuschadaptiven Vorverarbeitung von Sprachsignalen.

Die Wirksamkeit des Algorithmus wurde in subjektiven Hörversuchen mit acht normalhörenden Probanden validiert. Als Sprachmaterial diente der Oldenburger Satztest [2] (fester Sprachpegel von 60 dB SPL). Als Störgeräusch diente ein mit einem Kunstkopf in einem Kleinwagen aufgenommenes Fahrgeräusch bei 130 km/h, das seine primären Energieanteile im tiefem Frequenzbereich hatte (<125Hz). Der Pegel des Störgeräusches wurde zwischen 75 und 84 dB SPL variiert, um die gewünschten SNR zu erzeugen. Die Stimuli wurden über Kopfhörer (Sennheiser HD650) präsentiert.

Die untere dicke Linie in Abb. 2 zeigt die an die gemessenen Einzeldaten der Sprachverständlichkeit (Quadrate) angepasste psychometrische Funktion für die unverarbeiteten Sprachsignale. Die dargebotenen SNR von -18, -16 und -14 dB führten zu Worterkennungsraten von ca. 30, 50 und 80%. Die obere dicke Linie repräsentiert die psychometrische Funktion der durch AdaptDRC verarbeiteten Sprachsignale. Die Daten zeigen, dass selbst bei einem SNR von -24 dB im Mittel noch ca. 80% der dargebotenen Worte verstanden wurden. Bei -20 und -6 dB war die Verständlichkeit >90% bzw. nahezu optimal, d.h. es wurden erhebliche Gewinne durch den Algorithmus erzielt.



**Abbildung 2:** Sprachverständlichkeit von unverarbeiteten (untere dicke Linie), von durch AdaptDRC verarbeiteten (obere dicke Linie) und von durch einen Referenzalgorithmus nach [3] verarbeiteten Sprachsignalen.



**Abbildung 3:** Präferierte Frequenzgewichtung der Probanden für die dargebotenen Musikstücke. Die Kreisgröße entspricht der Häufigkeit der Präferenzwahl.

### Individualisierter Klang von Musik

Während der oben vorgestellte Algorithmus zur adaptiven Verständlichkeitsverbesserung vollständig autonom arbeitet, bedarf es bei der individuellen Anpassung von Equalizer- und Lautstärke-Einstellungen für Musikwiedergabe aufgrund der hohen Individualität der Präferenzen einer direkten Einstellung durch den Nutzer. Bestehende Möglichkeiten zur Klangindividualisierung in Wiedergabensystemen im Auto sind jedoch stark begrenzt und gehen häufig nicht über wählbare Presets (Jazz, Pop, etc.) bzw. relativ umständlich und für Laien schwer bedienbare Equalizer-Einstellungen hinaus. Ein weiteres Ziel dieser Studie war es daher, eine neuartige, intuitive und auch durch ältere Nutzer leicht zu bedienende [4] Nutzerschnittstelle zur Klangindividualisierung zu validieren und damit die zu erwartende interindividuelle Streuung von Klangpräferenzen zu untersuchen.

Die Nutzerschnittstelle wurde mit 15 normalhörenden Probanden getestet und bestand aus einer Kombination von einer 2D-Touch-Oberfläche zur Veränderung der Equalizer-Einstellungen und einem haptischen Drehknopf zur Veränderung der Gesamtlautstärke. Über die 2D-Oberfläche konnten die Probanden die Frequenzgewichtung von basslastig bis höhenlastig (19 Verstärkungspresets entlang der horizontalen Achse) und die Gewichtung des Mittenbereiches (19 Presets entlang der vertikalen Achse) verändern. Die 19x19 Presets waren so gewählt, dass eine kontinuierliche Bewegung auf der 2D-Oberfläche einer perzeptiv-kontinuierlichen Klangänderung entsprach, d.h. die Wechsel zwischen den Presets waren kaum bzw. gar nicht als Sprünge hörbar. Die Aufgabe der Probanden bestand darin, die Position auf der 2D-Oberfläche und die die Gesamtlautstärke zu finden, die der persönlichen Klangpräferenz entsprach. Als Stimuli dienten 26 verschiedene Musiksignale, die über Kopfhörer (HD650) dargeboten wurden. Die Gesamtheit der von allen Probanden ausgewählten Preset-Kombinationen ist in Abb. 3 dargestellt. Die kleinsten Kreise repräsentieren Einzelantworten, größere Kreise entsprechend mehrfach gewählte Datenpunkte.

Insgesamt zeigen die Daten eine erhebliche Streuung der präferierten Klangeinstellungen. Die wenigsten Datenpunkte lagen im Bereich der flachen Equalizer-Einstellung (Kreuz), d.h. fast immer wurde eine Anpassung des Klangs präferiert. Der einzige über die Probanden konsistente Trend war eine

Bevorzugung von Mittenverstärkung gegenüber Mittenabschwächung (Presets überwiegend  $\geq 10$ ). Bei der Bass-/Höhenverstärkung lagen die Präferenzen zwischen Preset 1 und 17, wobei eine leichte Tendenz zu einer präferierten Höhenbetonung (Presets 13-15) auftrat. Auch auf Basis einzelner Probanden wurden erhebliche Unterschiede beobachtet: während einige Probanden die Presets sehr konsistent über die verschiedenen Stimuli auswählten, nutzten andere Probanden fast den gesamten Preset-Raum für unterschiedliche Stimuli. Auch in Bezug auf die präferierte Gesamtverstärkung traten große interindividuelle Varianzen und Unterschiede von bis zu ca. 30 dB auf.

### Diskussion und Zusammenfassung

In Bezug auf die Präferenzen von Equalizer-Einstellungen zeigte diese Studie einen erheblichen Bedarf an Individualisierungsmöglichkeiten auf, der in bestehenden Automotivesystemen nicht adäquat adressiert wird. Da die Einflussfaktoren auf die großen interindividuellen Streuungen der bevorzugten Einstellungen bisher nicht verstanden und durch perzeptive Modelle vorhergesagt werden können [5], sind intuitive und leicht bedienbare Nutzerschnittstellen notwendig, wie z.B. das hier untersuchte 2D-Touch-Verfahren. Durch derartige Schnittstellen ist es auch möglich veränderte Klangpräferenzen durch Auftreten von sich verändernden Fahrgeräuschen nachzuregeln [6].

Für Wiedergabesysteme von Sprachsignalen (z.B. Autoradiosprecher, Navigationssystem) ist es hingegen möglich auch ohne direkte Nutzereinbindung eine erhebliche Verbesserung der Verständlichkeit zu erzielen. Der modellgesteuerte Algorithmus AdaptDRC passt hierbei das wiederzugebende Signal in Echtzeit adaptiv an vorliegende und zeitlich veränderliche Störgeräusch an, ohne dabei den Pegel zu erhöhen und dadurch ggf. eine unangenehm laute Wiedergabe zu bewirken.

### Literatur

- [1] Schepker, H., Rennie, J. und Doclo, S. (2013). Improving speech intelligibility in noise by SII-dependent preprocessing using frequency dependent amplification and dynamic range compression, Proc. Interspeech 2013, 3577-3581.
- [2] Wagener, K., Brand, T., Kollmeier, B. (1999). Entwicklung und Evaluation eines Satztests für die deutsche Sprache III: Evaluation des Oldenburger Satztests. Z. Audiol. 38, 86-95.
- [3] Sauert, B. und Vary, P. (2012). Near-end listening enhancement in the presence of bandpass noises, ITG Fachtagung Sprachkommunikation 2012.
- [4] Baumgartner, H., Paluch, R., Fuhrmann, K., Rennie, J., Meis, M. und Appell, J.-E. (2014). Usability evaluation of self-fitting interfaces for personalized sound systems. 17. Jahrestagung der DGA, 2014.
- [5] Saft, M., Hansen, M. und Rennie, J. (2014). Klangpräferenzen von Normalhörenden bei Musikwiedergabe. Fortschritte der Akustik - DAGA 2014.
- [6] Kubiak, A., Saft, M., Rennie, J. und Kollmeier, B. (2014). Listener's preferences with respect to frequency shaping and loudness adjustments for music and speech reproduction. Fortschritte der Akustik - DAGA 2014.