

Sprachqualität von Telefongesprächen aus dem Fahrzeug - ein noch immer ungelöstes Problem

Markus Lieb und Wilfried Gädicke

Infotainmententwicklung, Volkswagen AG, Wolfsburg, Email: Markus.Lieb, Wilfried.Gaedicke(at)volkswagen.de

Abstract

Im Gegensatz zur üblichen Mobil- oder VoIP-Telefonie hält das Telefonieren im Auto noch immer sehr große Herausforderungen bereit, um einen hohen Grad an empfundener Sprachqualität zu erreichen. Im Spannungsfeld zwischen Systemkosten und erzielbarer Leistungsfähigkeit werden beim Fahrzeughersteller verlässliche Testmethoden für Sprachübertragungsqualität benötigt, die über etablierte state-of-the-art Messverfahren hinausgehen. Mit dieser Arbeit zeigen wir die Faktoren, die die Sprachqualität im Fahrzeug negativ beeinflussen. Wir beschreiben die Schwierigkeiten bei der Beurteilung subjektiv empfundener Sprachqualität und deren Abbildung auf objektive Messverfahren. Dies führt zu fortschrittlichen Ansätzen in der Sprachqualitätsbewertung.

Wo liegen die Probleme ?

Bei der Telefonie im Fahrzeug unterscheidet man die beiden Übertragungsrichtungen "Uplink", aus dem Fahrzeug heraus, zum Beispiel ins Mobilfunk- oder Festnetz, und "Downlink", ins Fahrzeug hinein. Aufgrund der besonderen Bedingungen im Auto stellt vor allem der Uplink immer wieder eine große Herausforderung dar hinsichtlich einer guten Sprachqualität. Die im Fahrzeug störenden Einflussfaktoren ergeben sich aus der Freisprechsituation in der mobilen Umgebung:

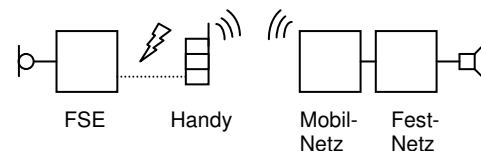
- Hintergrundgeräusche werden durch Motor und Lüftung, durch Fahrtwind und Reifenabrollgeräusche, sowie durch den umgebenden Verkehr verursacht.
- Die vergleichsweise große Distanz zwischen dem Freisprechmikrophon und dem Sprechermund, typischerweise 50cm, verschlechtert das im Fahrzeug ohnehin vorhandene niedrige Signal- zu Rausch-Verhältnis nochmals deutlich.
- Die in der Freisprechsituation aktive Downlink-Wiedergabe über die Fahrzeug-Lautsprecher verursacht ohne eine leistungsstarke Echokompensation störende Echos im Uplink.

Der Kommunikationspartner am entfernten Ende nimmt bei nicht optimalen Systemen eine Reihe von Qualitätseindrücken wahr. Sprache kann dumpf, schrill, verzerrt, metallisch, blechern, hohl oder rau klingen. Wenn Wortfragmente fehlen wirkt die Sprache abgehackt bis hin zu hörbaren Aussetzern. Wenn Hintergrundgeräusche zu stark bedämpft werden, also die Fahrsituation am entfernten Ende nicht mehr nachempfunden werden kann, wird dies als ebenso störend empfunden, wie zu starke Hintergrundgeräusche oder gar Echos der eigenen Stimme. Als besondere Qualitätseinbuße wird bewertet, wenn in der Double Talk Situation, also beide Sprecher gleichzeitig

aktiv, nicht beide Kanäle, Up- und Downlink, ausgewogen übertragen, sondern, im Extremfall, ein harter Umschaltvorgang zwischen den beiden Gesprächspartnern erfolgt.

Die lange Kette des Übertragungsweges

Folgendes Blockschaltbild verdeutlicht die typische Verarbeitungskette des Sprachsignals vom Mikrophon im Fahrzeug bis zum Lautsprecher am Hörer eines Festnetzanschlusses:



Die Freisprecheinrichtung (FSE) im Fahrzeug ist dabei mit dem Mobiltelefon entweder drahtlos via Bluetooth oder über proprietäre drahtgebundene Schnittstellen verbunden.

Die Einflussmöglichkeiten

Bei der Auslegung einer Freisprecheinrichtung spielen eine Reihe von Anforderungen eine Rolle:

- Bei der Platzierung des Freisprechmikrophons gilt es den ausgewogenen Kompromiss zu finden zwischen möglichst weiter Entfernung von Lüftungsdüsen, Scheiben und Lautsprechern einerseits und möglichst großer Nähe zum Sprechermund andererseits. Auch die Wahlmöglichkeit zwischen einer für den Fahrer optimierten oder einer für Fahrer und Beifahrer gleichermaßen ausgewogenen Lösung, und nicht zuletzt das Design ist bei der Mikrophon-Platzierung und Ausrichtung mitentscheidend.
- Die in der FSE eingesetzten Sprachverarbeitungs-algorithmen zur *Noise Reduction* und *Echo Cancellation* müssen sorgfältig auf die jeweilige Fahrzeug-Kabine und -Umgebung abgestimmt werden.
- Das Zusammenspiel fahrzeugseitiger Algorithmen und Verarbeitungsschritte, die im Handy vorgenommen werden, ist bei der Optimierung der Sprachqualität ein wichtiger Ansatzpunkt.

Für den Fahrzeughersteller ist es wichtig, sich von den vergleichsweise kurzen Produktzyklen der Handy-Hersteller zu entkoppeln und seine Freisprecheinrichtung universell, das heißt mit möglichst vielen - auch künftigen - Mobiltelefonen kompatibel, auszulegen. Dieses Ziel steht im Widerspruch zur Notwendigkeit, Optimierungen individuell für einzelne Handy-Typen vorzunehmen. So ist es beispielsweise bei manchen Handys möglich, interne

Verfahren zur Noise Reduction und Echo Cancellation im Freisprechfall abzuschalten und somit die vollständige Kontrolle durch die FSE zu ermöglichen. Andere Handys erlauben solche Maßnahmen nicht, oder nur unvollständig, was zu einer großen Anzahl an proprietären Optimierungen führt.

Keinen Einfluss hat der FSE-Hersteller auf die zahlreichen netzseitigen Verarbeitungsschritte, wie den Einsatz von Sprach- Codecs, Frequenzgangs-Änderungen, Pegelwaagen, DTX (*Discontinuous Transmission*), Einspeisung von *Comfort Noise*, Übertragungs-Problemen, wie *Burst-Errors*, die sich in kurzzeitigen *Drop-Outs* der übertragenen Sprache äußern können.

Als Beispiel: In einer Kurzstudie wurden bei Volkswagen Untersuchungen zur Lautstärke in unterschiedlichen Übertragungssituationen angestellt. Variiert wurden dabei ausschließlich netzrelevante Eigenschaften, wie Tageszeit, Ort, Netzbetreiber, nicht aber Fahrzeug- oder Telefon-Konfiguration, Testaufbau oder Testsignale.

Die mit *artificial speech* Signalen gemäß P.50 ermittelten Lautstärken zeigten statistisch annähernd eine Gleichverteilung mit Streubreiten von 12dB zwischen leisesten und lautesten ermittelten Pegeln. Diese große Pegelstreuung ist durchaus problematisch und kann zu Qualitätsbeanstandungen führen ("Lautstärke aus dem Fahrzeug heraus zu gering").

Kosten vs. Performance

Neben den dargestellten Themen bei der Auslegung einer FSE kann fahrzeugseitig weiteres Verbesserungspotential erschlossen werden. Vor dem Einsatz solcher Maßnahmen muss abgewogen werden zwischen der Erhöhung des Aufwands, also der Systemkosten, und dem erzielbaren Performance-Gewinn. Soll ein Mikrofon-Array eingesetzt werden, so steht ein verbessertes Signal-zu-Geräusch-Verhältnis dem erhöhten Aufwand für den Verbau mehrerer Mikrophone sowie die zusätzliche Signalverarbeitung gegenüber. Dieser Verarbeitungsaufwand ist bei einem dynamischen Array mit adaptivem *Beam-Forming* entsprechend höher als bei einem statisch ausgelegten System mit zum Beispiel fixer Fokussierung auf den Fahrer. Mehrere Mikrophone können neben klassischer Verwendung in einer Array-Struktur auch in Form von *Noise Reference Acquisition* eingesetzt werden, bei der einzelne Mikrophone die Hintergrundgeräusche möglichst ohne Sprachsignalanteile erfassen und somit eine zusätzliche, bessere, Referenz für die Noise Reduction Algorithmen liefern. Ebenfalls zum Themenbereich "Mikrophon" zählen neuartige Platzierungen, wie beispielsweise an den Kopfstützen oder integriert im Sicherheitsgurt.

Eine vielversprechende Möglichkeit, die oben erwähnte Problematik von proprietären Handy-Optimierungen in der Sprachverarbeitung zu vermeiden, ist der Einsatz von Festeinbau-Telefonen, oder als jüngste Entwicklung, von Handy-Vorbereitungen basierend auf dem Bluetooth Profil RSAP (*remote SIM access profile*). Hierbei fungiert das Handy als "drahtloser SIM-Karten-Leser", die FSE enthält die komplette GSM-Verarbeitung. Ein solches System

eliminiert einige negative Einflussfaktoren, die aus der Koppelung von Freisprecheinrichtung und Mobiltelefon im herkömmlichen Setup entstehen, dort vor allem verursacht durch kaskadierte Signalverarbeitungs-Algorithmen im Fahrzeug und Handy und begründet durch die höchst unterschiedlich arbeitenden Verarbeitungssysteme in Telefonen unterschiedlicher Hersteller.

Test- und Messverfahren

Das übliche Prozedere bei der Beurteilung von Systemen hinsichtlich ihrer Sprachqualität besteht aus aufwändigen Benutzertests. Die Schwierigkeiten bei der Beurteilung subjektiv empfundener Sprachqualität und bei deren Abbildung auf objektive Messverfahren beschäftigen Fahrzeug- wie Meßequipment-Hersteller. So arbeitet ein VDA-Arbeitskreis an einer allgemeinen Spezifikation [5]. Darin werden zahlreiche Messungen technischer Parameter, wie Frequenzgänge, Pegel, Latenzzeiten, Double Talk Verhalten, sowie reproduzierbare Messverfahren festlegt. Ziel sind objektive, vergleichbare Aussage über die Eigenschaften und Qualitäten unterschiedlicher Systeme.

Die Erfahrungen zeigen, dass ein Einhalten der bisher festgelegten Spezifikationen zwar notwendig aber noch nicht hinreichend für eine subjektiv als gut empfundene Sprachqualität ist. Psychoakustisch motivierte Messverfahren wie TOSQA oder PESQ zielen in die richtige Richtung, nämlich "weiche" Faktoren der Sprachqualität wie "abgehackter Klang" mit zu berücksichtigen bei der technischen Erfassung von Messgrößen. Ein breiter Einsatz von MOS- (*mean opinion score*) Werten zu einzelnen Aspekten und Parametern der Sprachqualität ist hier die Entwicklungstendenz. Hauptschwierigkeit bleibt die Korrelation zwischen subjektiv ermittelten Werten und technisch erfassten und errechneten Kenngrößen zu erhöhen.

Was die "richtigen" Fragen und Kriterien sind, muß dabei ganz zu Anfang jeder neuen Algorithmus- und Verfahrensentwicklung stehen.

Vom zweiten Ziel, nach einer vergleichenden Beurteilung, aus Messergebnissen auch direkte eindeutige Rückschlüsse auf Optimierungen an der FSE zu erhalten ist man noch ein gutes Stück entfernt.

Literatur

[1] Blauert, J., "Communication Acoustics", Springer, Berlin, 2005.

[2] Gay, S. and Benesty, J., "Acoustic Signal Processing for Telecommunication", Kluwer, Boston, 2001

[3] Reference to Head Acoustics homepage.
URL: <http://www.head-acoustics.de>

[4] Reference to Brüel & Kjaer homepage.
URL: <http://www.bruelkjaer.de>

[5] Kettler, F., "Test- Spezifikation", in Funkschau, 24/2002