

Modellierung der Einflüsse des Übertragungskanals bei der Interaktion mit einem Smart-Home-System

Jan Krebber, Alexander Raake, Sebastian Möller

Institut für Kommunikationsakustik (IKA), Ruhr-Universität Bochum, D-44780 Bochum

krebber@ika.ruhr-uni-bochum.de, raake@ika.ruhr-uni-bochum.de, moeller@ika.ruhr-uni-bochum.de

1. Einleitung

Im Rahmen des europäischen IST-Projektes INSPIRE (INfotainment management with Speech Interaction via REmote microphones and telephone interfaces) [1] wird ein Sprachdialogsystem zur Interaktion mit verschiedenen Hausgeräten (TV, Videorecorder, Lampen, Jalousien, etc.) erstellt. Zweck dieses Dialogsystems ist es, ein intelligentes und einheitliches Benutzerinterface bereit zu stellen, welches die komplizierte Interaktion mit unterschiedlichen Geräten erleichtert und somit die Qualität der Akzeptanz der Dienste erhöht. Das System ist mit der Außenwelt sowohl über das Telefonnetz als auch über das Internet verbunden. Dadurch ist es möglich, Informationen aus dem Internet abzufragen und das System von einem entfernten Ort aus anzurufen.

Beim Sprachzugang über das Telefonnetz oder über das Internet (Voice over Internet Protocol, VoIP) treten jedoch Störungen auf, die die Interaktion mit dem System maßgeblich beeinflussen können. Daher sollen diese schon in der Planungsphase berücksichtigt und ihr Einfluss auf die Systemqualität minimiert werden. Hierzu wurde ein Simulationssystem erstellt, welches sowohl stationäre Störungen in traditionellen Netzen als auch für VoIP typische zeitlich instationäre Störungen kontrolliert generieren kann. Das Simulationssystem wurde um eine Autotelefon-Simulation erweitert, um möglichst viele Benutzungsszenarien ermöglichen zu können. Es wird gezeigt, wie die Simulation wirkungsvoll zur Optimierung der Komponenten des Sprachdialogsystems eingesetzt werden kann.

2. Warum Simulation?

Die Simulation einer Telefonverbindung bietet gegenüber der Verwendung realer Telefonverbindungen eine Vielzahl von Vorteilen. Reale Netzwerke lassen nur geringe oder keine Kontrolle über die Charakteristik oder die Parameter der Verbindung zu. Außerdem sind relevante Netzwerkdaten der jeweiligen Verbindung nicht verfügbar. Somit sind keine gesicherten Aussagen über eventuelle Schwächen einzelner Komponenten möglich. Dies ist aber notwendig, um einen Spracherkenner gezielt gegen Übertragungseinflüsse robust zu machen.

Wenn keine Parameterwerte über die Verbindung zur Verfügung stehen, muss eine sehr große Anzahl von Testdaten gesammelt werden, um möglichst alle Verbindungseinflüsse repräsentativ abzudecken. Dies ist zeit- und kostenintensiv. Ein Simulationssystem ermöglicht hier ein sehr viel kleineres Korpus (Faktor 50), weil die Übertragungseigenschaften der Verbindung bekannt sind. Es werden keine ungewollt redundanten Verbindungen verwendet.

Verwendet man aufgezeichnete „saubere“ Sprachdaten am Eingang der Simulation, so können durch einfache Veränderung der Verbindungsparameter große Mengen Trainingsmaterials für Sprach- und Spracherkenner erzeugt werden.

3. Aufbau der „allgemeinen“ Telefonsimulation

Die Simulation von PSTN, ISDN und VoIP Verbindungen erfolgt nach der ITU-T Recommendation G.107 [2] [Fig. 1]. Das Modell beinhaltet alle Parameter, die für eine Netzwerkplanung notwendig sind. Mit diesen Referenzmodellen können analoge 2-Draht/4-Draht-

Verbindungen oder digitale Netzwerke, die mit Handapparaten als Endgeräten arbeiten, beschrieben werden.

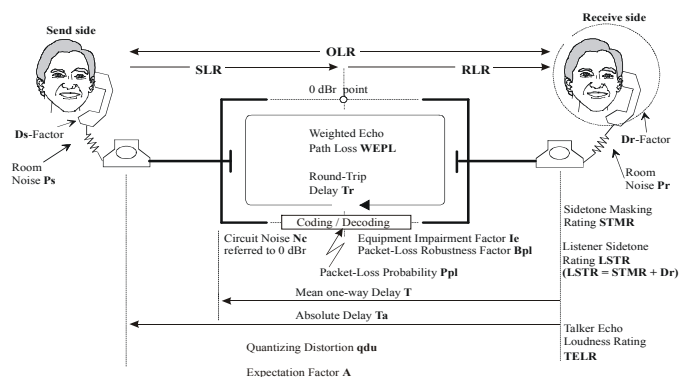


Fig. 1: Parametrische Beschreibung des Übertragungskanals nach ITU-T Rec. G.107.

Neben dem reinen Sprachübertragungspfad durch das Netzwerk werden auch die Sprecher- und Hörerechos sowie die elektrische Kopplung zwischen Sprechermikrofon und Sprecherhörerkapsel berücksichtigt. Das hier verwendete Referenzmodell bezieht sich auf synchrone Telefonnetze, an denen normale Handapparate angeschlossen sind. Dieses Modell kann einfach auf weitere Szenarien erweitert werden. Möchte man z.B. Headsets oder Freisprecher anstelle der Handapparate verwenden, so braucht man hierzu lediglich die Übertragungsfunktion zwischen Mensch und elektrischer Netzwerkanbindung zu verwenden. Methoden zur Bestimmung dieser akustisch-elektrischen Übertragungsfunktion sind in den ITU-Empfehlungen der P.20- und der P.300-Serie beschrieben. Die zeitvarianten Störungen, die in Mobilfunknetzen und der VoIP-Telephonie vorkommen, haben einen erheblichen Einfluss auf die Übertragungsqualität der Sprache und damit auch auf die Qualität des Dialogsystems und die damit verbundene Akzeptanz des sprachbasierten Fernzugangs zu Smart-Home-Systemen. Im Referenzmodell werden diese Störungen vereinfachend durch die Wahrscheinlichkeit von Paketverlusten und durch einen Faktor, der die Robustheit gegen Paketverluste widerspiegelt, beschrieben. Dieses Modell wird nun in ein Blockdiagramm überführt [3] [Fig. 2].

Die Bezeichnungen im Blockdiagramm haben folgende Bedeutungen:

T...	→ Verzögerungsstufe
BP	→ Bandpass
Codec	→ Kodier-/Dekodieralgorithmus
Lst/Le	→ Listener Sidetone/Listener Echo
WEPL	→ Weighted Echo Loss Path
SLR/RLR	→ Sender Loudness Ratio/Receiver L R
Nc	→ Circuit noise (Rauschen vor der Kodierung)
Nfor	→ Rauschen im Handapparat und Teilnehmeranschluß (Empfängerseite)

An das Simulationssystem können unterschiedliche Endgeräte im 4-Drahtmodus angeschlossen werden (im Gegensatz zu Fig. 1, in welcher 2-Draht-4-Draht-Übergänge angenommen werden). Die elektroakustischen Eigenschaften der einzelnen Endgeräte (SLR_{set} und RLR_{set}) sind mit einem sprechenden Kunstkopf vermessen worden [5]. Die resultierende Übertragungsfunktionen werden dann über

SLR' und RLR' an die von der ITU vorgegebenen Frequenzgänge eines „intermediate reference systems“ [4] angepaßt.

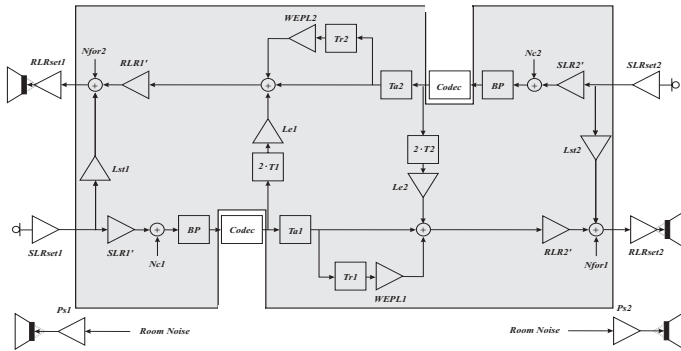


Fig. 2: Blockdiagramm einer PSTN/ISDN/VoIP Simulation [3].

Der grauhinterlegte Bereich der Simulation ist auf einer DSP-Farm implementiert. Die Coders im Übertragungspfad werden auf externen Codec-Karten berechnet. Dazu muss das Signal aus der DSP-Farm über A/D und D/A Wandler herausgeführt werden. An diesen Schnittstellen wird auch die VoIP-Simulation eingefügt, d.h., dass die VoIP-Simulation analoge Schnittstellen aufweisen muss. Die VoIP-Simulation wurde wie in Fig. 3 realisiert.

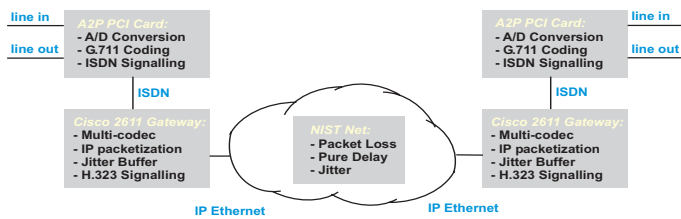


Fig. 3 : VoIP-Simulation.

Das analoge Line-Signal wird A/D gewandelt nach G.711 [2] (logarithmische PCM) kodiert und auf ISDN-Ebene an ein Gateway übertragen. Im Gateway findet eine Transcodierung hin zum gewünschten IP-Übertragungscodec statt. Anschließend wird das Signal in Pakete unterteilt, mit einem IP-Header versehen und über die Ethernetschnittstelle verschickt. Das Internet wird durch die Software NIST-Net [5] simuliert, welche auf einem separaten Rechner läuft. Diese Software ist in der Lage, Paketverluste (packet loss), Verzögerung (delay) und Jitter (Pakete mit unterschiedlicher Übertragungslaufzeit) zu generieren. Damit wird die variierende Anzahl an Routern und ein unterschiedliches Datenverkehrsaufkommen simuliert. Anschließend gelangen die IP-Pakete zum zweiten Router. Dort werden sie im Jitterbuffer zwischengespeichert. Dieser gleicht die unterschiedlichen Paketlaufzeiten aus und sortiert die Pakete, gemäß ihrer Position im ursprünglichen Sprachsignal. Anschließend findet eine Transcodierung nach G.711 und die Ausgabe als ISDN-Signal statt. Zuletzt wird das Signal D/A-gewandelt und analog ausgegeben.

4. Aufbau der Auto-Freisprech-Simulation

Wie in Fig. 4 zu sehen wird die PSTN/ISDN/VoIP-Simulation durch die CAR-Simulation erweitert. Dabei werden folgende Übertragungswege und Geräuschquellen berücksichtigt:

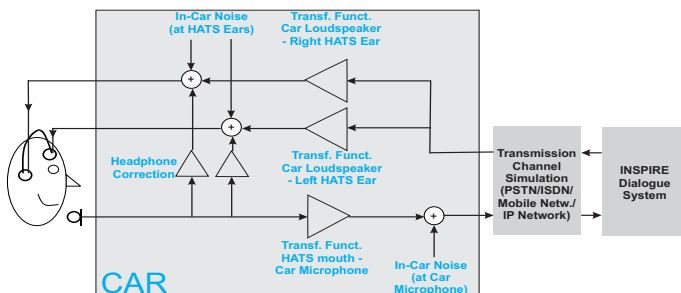


Fig. 4: Car-Simulation als Erweiterung der PSTN/ISDN/VoIP-Simulation.

- Außengeräusche bei verschiedenen konstanten Geschwindigkeiten, die vom Benutzer wahrgenommen werden
- Außengeräusche bei verschiedenen konstanten Geschwindigkeiten, die auf das Fahrgastraummikrofon einwirken
- Übertragungsstrecke Frontlautsprecher → Benutzer
- Übertragungsstrecke Benutzer → Fahrgastraummikrofon
- Kopfhörerentzerrung bezüglich des eigenen Sprachsignals

Die CAR-Simulation umfasst momentan lediglich eine akustische Testumgebung und keine weiteren multimodalen Elemente.

5. Anwendungen

Wie in Fig. 5 zu sehen, können zur Evaluierung des Dialogsystems nun Sprachdaten über ein normales, drahtgebundenes Telefon (Handset), ein Funktelefon (Mobile Phone), ein Headset oder einen Freisprecher (Hands Free Terminal, HFT) in die Simulation gegeben. Diese werden anschließend an das Dialogsystem weitergeleitet. Dort werden die Sprachdaten verarbeitet und die akustische Ausgabe des Dialogsystems über die Simulation an den Benutzer ausgegeben. Um einen großen Korpus an Trainingsdaten zu gewinnen, zeichnet man die Sprachdaten noch vor der Simulation auf, verarbeitet die aufgenommenen Sprachdaten mit den gewünschten Beeinträchtigungen und nimmt diese danach wieder auf. Somit hat man während der Evaluationsphase immer identisches Trainingsmaterial zur Verfügung. Darüber hinaus können auch später noch Beeinträchtigungen berücksichtigt werden, die man während der Planung noch nicht beachtet hat.

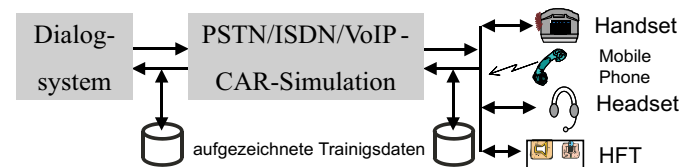


Fig. 5: PSTN/ISDN/VoIP-Simulation als Bestandteil der Evaluierung eines Dialogsystems.

6. Diskussion und Ausblick

Das PSTN/ISDN/VoIP-Simulationssystem ist in der Lage, normale Telefonverbindungen und Autofreisprechszenarien zu simulieren. Dabei können alle bislang bekannten Parameter und Beeinträchtigungen einer Fernsprechübertragung eingestellt und gezielt verändert werden. Sprecher- und Spracherkennungssysteme können somit wohldefiniert trainiert und getestet werden. Darüber hinaus können genaue Angaben zur Erkennungsleistung in Abhängigkeit einzelner Übertragungsparameter gewonnen werden. Ebenso lässt sich das Sprachverständnis synthetisch erzeugter Sprache überprüfen, die über den Telefonkanal ausgegeben wird. Des Weiteren ist es möglich, die Einflüsse des Übertragungskanal in realistischen Kommunikationssituationen zu evaluieren.

Literatur

[1] INSPIRE (2003), www.inspire-project.org
 [2] ITU-T Rec. G.711 (1988). Pulse Code Modulation of Voice Frequencies. International Telecommunication Union, CH-Genf.
 [3] Möller, S. (2000). Assessment and Prediction of Speech Quality in Telecommunications. Kluwer Academic Press, USA-Boston.
 [4] ITU-T Rec. P.48 (1988). Specifications for an Intermediate Reference System (Blue Book Edition). International Telecommunication Union, CH-Genf
 [5] National Institute of Standards and Technology (2002). NIST Net. www.nist.gov
 [] Möller S. and Raake, A. (2002). Telephone Speech Quality Prediction: Towards Network Planing and Monitoring Models for Modem Network scenario. Speech Communication, 38, 47-45.