VON DER SCHMALBAND-TELEFONIE ZUR MOBILEN AUDIO-KOMMUNIKATION

Peter Vary
Institut für Nachrichtengeräte & DV, RWTH Aachen University
vary@rwth-aachen.de

Zusammenfassung: Sprachtelefonie wird voraussichtlich die dominierende Nutzungsform des Mobiltelefons bleiben. Da die Audio-Frequenzbandbreite noch immer auf ca. 3.4 kHz begrenzt wird, besteht ein sehr hohes Potential zur Qualitätsverbesserung. Es werden aktuelle Entwicklung der Sprach-Audio-Codierung sowie unterstützende Maßnahmen zur künstlichen Bandbreiten-Erweiterung vorgestellt. Eine besonders interessante Variante ist die Bandbreiten-Erweiterung, die zusätzliche Bits zur Qualitätsverbesserung nutzt, die mit Methoden der Steganographie (digitale Wasserzeichentechnik) im Schmalband-Datenstrom übertragen werden. Weiterhin wird gezeigt, dass mit iterativer Quellen-Kanaldecodierung die verbesserte Audio-Qualität auch über stark gestörte Mobilfunkkanäle gewährleistet werden kann.

1. Was lässt sich an modernen Mobiltelefonen noch verbessern?

Luxus-Telefone mit Gold und Diamanten in der Preisklasse von Rolex Armbanduhren sind sicher eine Möglichkeit, die aber nur für einen kleinen exklusiven Kundenkreis von Interesse sein dürfte. Aus Ingenieurssicht stellt sich eher die Frage: Warum klingt ein (Mobil-)Telefon wie ein Telefon aus hoheitlichen Bundespostzeiten – auch „Plain Old Telephone“ (POT) genannt. Während Sprache einen natürlichen Frequenzumfang von rund 10 kHz hat wurde zu Zeiten des POTs die Frequenzbandbreite auf 3.4 kHz begrenzt, wodurch einerseits eine als ausreichend eingeschätzte Silbenverständlichkeit erreicht wurde und andererseits eine effiziente Weitverkehrstelefonie mit Frequenzmultiplexübertragung auf Koax-Kabeln und Richtfunkstrecken entwickelt werden konnte. Diese traditionelle Bandbreitenbegrenzung wurde aus Kompatibilitätsgründen auch im ISDN und in den GSM-Mobilfunknetzen beibehalten.

2. Die Entwicklung der Mobilfunknetzte

Die verfügbaren Bitraten in den „Next Generation Mobile Networks“ der nächsten Jahre werden dramatisch ansteigen, so dass die tradierte Limitierung nicht mehr notwendig erscheint. Schon heute können mit UMTS HSPA („High Speed Packet Access“) Bitraten von mehr als 2 MBit/s erreicht werden, UMTS Long-Term-Evolution („LTE“) wird diese Bitraten noch vervielfachen. Dennoch wird in vielen Regionen aus (Infrastruktur-) Kostengründen nur das vergleichsweise langsame „GSM/EDGE“ als „Fall-Back“ Lösung die flächendeckende Mobilfunk-Versorgung bieten.

Um die Audio-Qualität nachhaltig zu verbessern ist es deshalb sinnvoll nicht nur auf höhere Bitraten zu setzen, sondern auch mit intelligenter Signalverarbeitung die Qualität unter Beibehaltung der Datenrate zu verbessern. Die verfügbare Rechenleistung wird weiter ansteigen, so dass Algorithmen, die vor wenigen Jahren noch nicht denkbar waren, kostengünstig realisiert werden können.

3. Das Prinzip der Modell-basierten Sprachübertragung

Die Sprachübertragung in den Mobilfunknetzen basiert auf einem Modell der Spracherzeugung. Im Prinzip werden, wie in Bild 1 skizziert, sendeseitig die Parameter

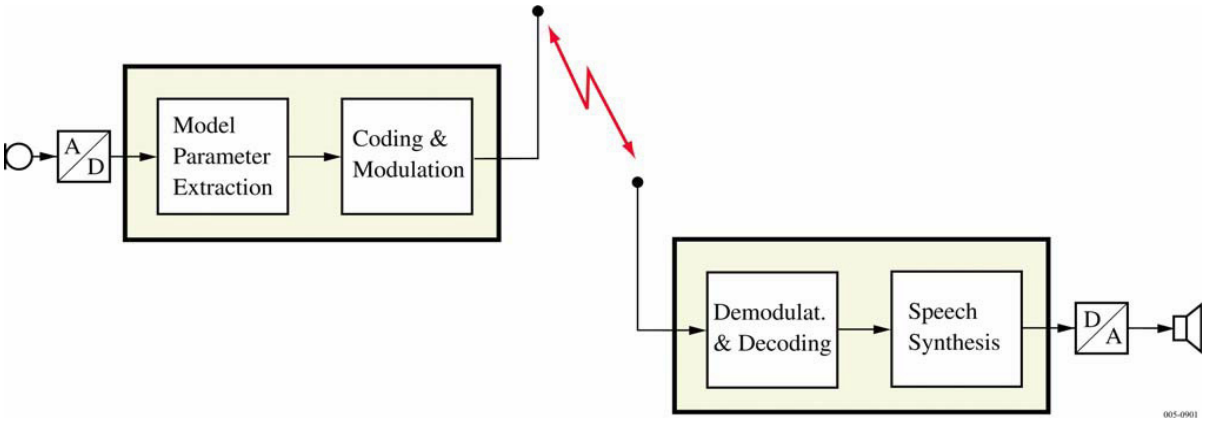

Bild 1: Prinzip der Modell-basierte Sprachübertragung über den Funkkanal

dieses Modells extrahiert und in codierter Form über die Funkstrecke übertragen. Auf der Empfangsseite wird die Sprache schließlich auf der Grundlage dieses Modells synthetisiert. Es handelt sich also um einen Vocoder, auch Sprachcodec genannt, mit natürlicher Sprachqualität.

Die Sprachqualitätsstufen der Sprachübertragung werden je nach Frequenzbandbreite in die folgenden Kategorien eingeteilt:

- Narrowband „NB“: 3.4 kHz
- Wideband „WB“: 7 kHz
- Super Wideband „SWB“: 14 kHz
- Fullband „FB“: 20 kHz

4. Die Adaptive Multi-Raten Codierung (AMR)

Für eine effiziente Sprachübertragung wird in den GSM Mobilfunksystemen als neuestes Verfahren die sog. Adaptive Multi-Rate Speech Codierung (AMR) eingesetzt, die im Wesentlichen die Robustheit gegenüber Bitfehlern infolge von Störungen auf dem Funkkanal verbessert (e.g. [1], [2]). Zur Fehlerverminderung wird die pro Kanal zur Verfügung stehende Bitrate von 22.8 kbit/s je nach Qualität der Funkverbindung mit 8 verschiedenen Modi dynamisch zwischen Sprache und Fehlerschutzbits aufgeteilt. Dadurch kann auch bei sehr schlechten Verbindungen eine ausreichende Schmalband-Sprachqualität (3.4 kHz) sichergestellt werden.

Der AMR-Codec beruht auf dem sog. CELP-Prinzip (Code Excited Linear Prediction), das in Bild 2 für die Empfangsseite dargestellt wird. Sendeseitig wird eine Vektorquantisierung des Anregungssignals für das zweistufige Synthesefilters nach dem Analyse-durch-Synthese Verfahren durchgeführt. Über den Funkkanal werden folgende Parameter übertragen:

- Vektor-Adressen (alle 5 ms)
- Verstärkungsfaktoren g (alle 5 ms)
- Filter-Koeffizienten P (Pitchfilter, alle 5 ms)
- Vokaltraktfilter A (alle 20 ms)

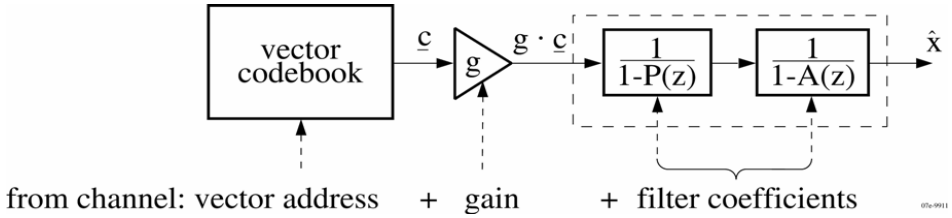

Bild 2: CELP Sprachdecoder (Code Excited Linear Prediction)

In den letzten zehn Jahren wurden verschiedene Versionen von AMR Codec-Standards sowohl für die Schmalbandtelefonie (3.4 kHz, [2]) als auch für die Breitbandtelefonie (7 kHz, [3]) entwickelt:

- AMR-NB: narrowband (ETSI, 1998), 3.4 kHz, fs=8 kHz, 4.75 … 12.2 kbit/s
- AMR-WB: wideband (ETSI/3GPP 2001), 7.0 kHz, fs=16 kHz, 6.6 … 23.85 kbit/s

5. Die Einführung von AMR-WB Codecs

In den Mobilfunknetzen ist bisher jedoch nur die Schmalbandversion verfügbar, während in der leitungsgebundenen Internet-Telefonie bereits zum Teil breitbandige Sprachcodierung eingesetzt wird. Der Mobilfunkbetreiber T-Mobile hat kürzlich in einem Feldtest in Hamburg und Köln den AMR-WB Codec für 150 Kunden mit speziellen UMTS-Telefonen verfügbar gemacht. Der Netzbetreiber hat angekündigt, dass dieser Codec mit der vergrößerten Audiobandbreite voraussichtlich in 2009/10 in einigen Netzbereichen eingeführt wird. Weitere Netzbetreiber haben ebenfalls großes Interesse an diesem Codec gezeigt.

Um diesen verbesserten Codec nutzen zu können, sind Investitionen in die Netze erforderlich. Insbesondere muss auf den Basisstationen neue Software für die Kanalcodierung und in den Mobilfunkvermittlungszentren neue Sprachcodecs installiert werden. Daneben sind die Protokolle entsprechend zu erweitern.

Vor allem müssen genügend viele Endgeräte mit dem neuen Codec zur Verfügung stehen. Die ersten Geräte, die WB-Codecs unterstützten werden derzeit von den Herstellern in den Markt gebracht, noch ist die Verbesserung in den Datenblättern gut versteckt (z.B. Ankündigung des neuen Nokia 6700 classic, Sony Ericsson C 905, unter Sprach-Codecs findet man WB-AMR, NB-AMR).

Neben der Ertüchtigung der Netze ist natürlich auch das Kompatibilitätsproblem zu lösen, damit auch die modernen „Terminals“ mit herkömmlichen Schmalbandtelefonen kommunizieren können. In der ersten Phase werden die „passenden“ WB-Endgeräte noch nicht weit verbreitet sein, so dass die neue Gesprächsqualität nicht oft zur Geltung kommt.

6. Bandbreiten-Erweiterung zur besseren Akzeptanz neuer Geräte

Die Akzeptanz der neuen Endgeräte kann durch empfangsseitige Maßnahmen zur künstlichen Bandbreiten-Erweiterung erheblich gesteigert werden (Bild 3). Derart aufgerüstete Mobiltelefone bieten selbst dann eine höhere Audioqualität wenn das Netz noch nicht umgerüstet ist, oder am anderen Ende der Verbindung noch ein herkömmliches schmalbandiges Telefon verwendet wird.

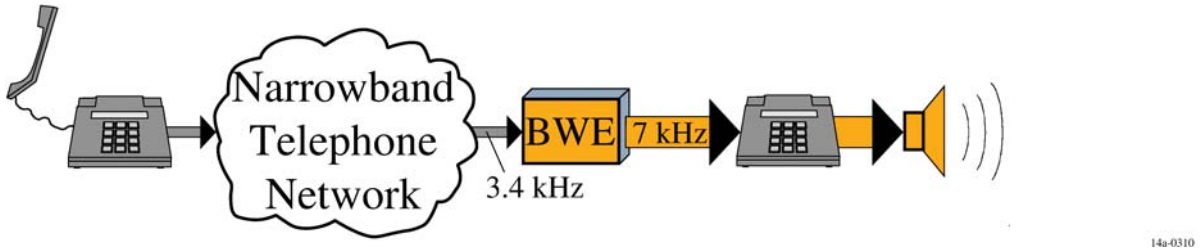

Bild 3: Künstliche Bandbreiten-Erweiterung durch empfangsseitige Signalverarbeitung

Die Bandbreiten-Erweiterung (BWE) nutzt Redundanz des Modells der Spracherzeugung und gewisse psychoakustische Eigenschaften des Gehörs. Mittels Merkmalserkennung und statistischer Schätzung werden aus dem schmalbandigen Sprachsignale die Parameter zur Synthetisierung der höheren Frequenzkoponenten (3.4 … 7.0 kHz) ermittelt. Es handelt sich im weitesten Sinne um eine spezielle Kombination von Spracherkennung und Sprachsynthese (e.g. [4]). Einen Eindruck von der Komplexität der zugrunde liegenden Theorie gibt Bild 4.

$$\tilde{\mathbf{y}}_{\mathrm{wb}} = \sum_{i=1}^{N_S} \mathrm{E}\{\mathbf{y}(m) \mid S_i(m)\} \cdot P(S_i(m) \mid \mathbf{X}(m_{\mathrm{k}}))$$

$$\mathrm{E}\{\mathbf{y}(m) \mid S_i(m), \mathbf{x}(m)\} = \sum_{l=1}^{L} \rho_{\mathbf{y}|\mathbf{x},il} \left(\boldsymbol{\mu}_{\mathbf{y},il} - \left((\mathbf{x} - \boldsymbol{\mu}_{\mathbf{x},il})^{\mathrm{T}} \mathbf{A}_{\mathbf{yx},il} \mathbf{A}_{\mathbf{yy},il}^{-1} \right)^{\mathrm{T}} \right)$$

$$P(S_i(m) \mid \mathbf{X}(m_{\mathrm{k}})) = \frac{\alpha_i(m)\, p(\mathbf{x}(m) \mid S_i(m))}{\sum_{i=1}^{N_S} \alpha_i(m)\, p(\mathbf{x}(m) \mid S_i(m))}$$

$$p(\mathbf{x}(m) \mid S_i(m)) = \sum_{l=1}^{L} \rho_{\mathbf{x},il} \frac{\sqrt{\det \mathbf{A}_{\mathbf{x},il}}}{(2\pi)^{d/2}} \exp\left(-\frac{1}{2} (\mathbf{y} - \boldsymbol{\mu}_{\mathbf{x},il})^{\mathrm{T}} \mathbf{A}_{\mathbf{x},il} (\mathbf{y} - \boldsymbol{\mu}_{\mathbf{x},il}) \right)$$

$$\alpha_i(m+1) = \sum_{j=1}^{N_S} \alpha_j(m)\, p(\mathbf{x}(m) \mid S_j(m)) P(S_i(m+1) \mid S_j(m)), \quad \alpha_i(1) = P(S_i)$$

Bild 4: Zur Theorie der künstlichen Bandbreiten-Erweiterung von Telefonsprache

7. Bandbreiten-Erweiterung mit versteckter Nebeninformation

Ein darüber hinausgehender Ansatz, die Zeit bis zur vollständigen Umrüstung der Mobilfunknetze zu überbrücken besteht darin, zusätzliche Bits zur Signalverbesserung mit Methoden der Steganographie (digitale Wasserzeichentechnik) im Schmalband-Datenstrom förmlich zu verstecken (bis zu 2 kbit/s im AMR-NB Codec, [5]). In Bild 5 wird angedeutet, dass dies beim Teilnehmer A geschieht. Der so modifizierte Datenstrom kann von einem herkömmlichen AMR Decoder (Teilnehmer B in Bild 5) ohne Beeinträchtigung der Sprachqualität decodiert werden kann. Der große Vorteil dieses Ansatzes ist, dass keine Veränderungen am Netz erforderlich sind. Wenn das codierte Sprachsignal ohne Umcodierung vom Teilnehmer A zum Teilnehmer C transportiert wird, der über ein entsprechend aufgerüstetes Terminal verfügt, kann dort mittels der versteckten Information das breitbandige Sprachsignal decodieren werden.

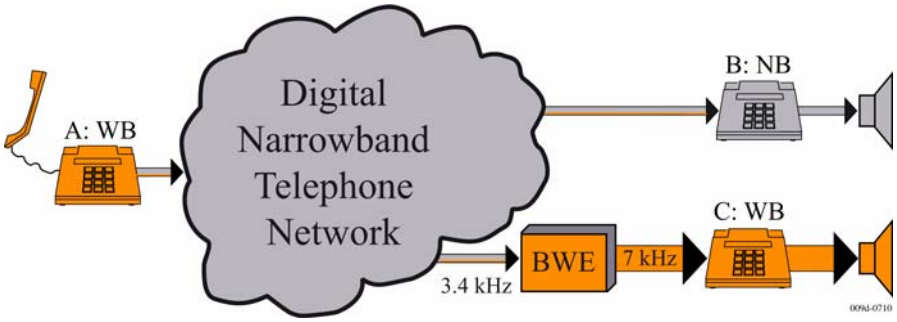

Bild 5: Kompatible Bandbreitenerweiterung mit versteckter Nebeninformation

8. Iterative Fehlerkorrektur und Fehlerverdeckung

Die durch neue Codecs oder durch Bandbreitenerweiterung mit Nebeninformation erzielbare bessere Audio-Qualität kann auch über stark gestörte Mobilfunkkanäle gewährleistet werden, insbesondere durch den Fehlerschutz mittels iterativer Quellen-Kanaldecodierung. Das aus dem Bereich der Kanalcodierung bekannte Turbo-Verfahren kann, wie in Bild 6 skizziert, auf die Kombination von Kanaldecoder und Sprachdecoder erweitert werden. Der Sprachdecoder liefert dabei aufgrund von Restredundanz der Modellparameter extrinsische Information für die iterative Kanalcodierung auf Bitebene (Turbo-Error-Concealment, e.g. [6]).

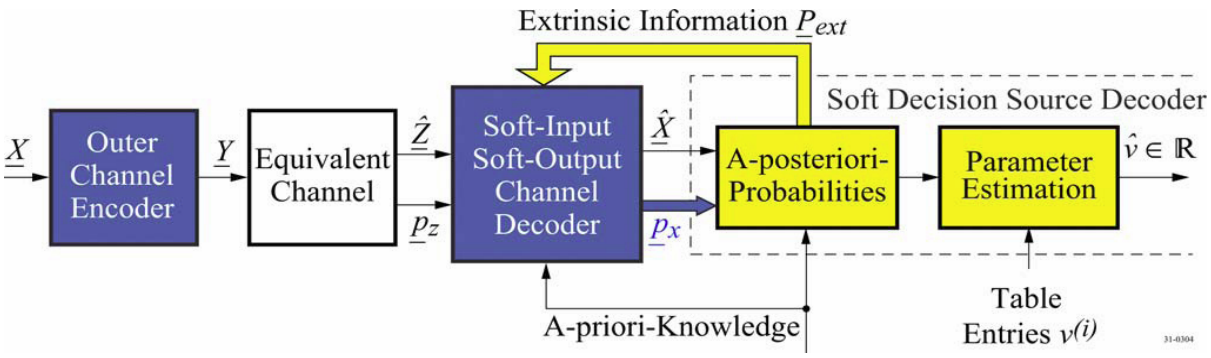

Bild 6: Iterative Quellen-Kanaldecodierung

9. Ausblick: Binaurale Telefonie

Die Audio-Qualität von Mobiltelefonen wird voraussichtlich in wenigen Jahren die Qualität der Festnetztelefonie weit hinter sich lassen. Der nächste Schritt könnte dann die binaurale Telefonie sein, die räumliches Hören über Mobilfunkverbindungen ermöglicht, so als säße man mit am Konferenztisch (Bild7). Geeignete Verfahren zur Stereo Codierung stehen bereits heute zur Verfügung (e.g. [7], [8]).

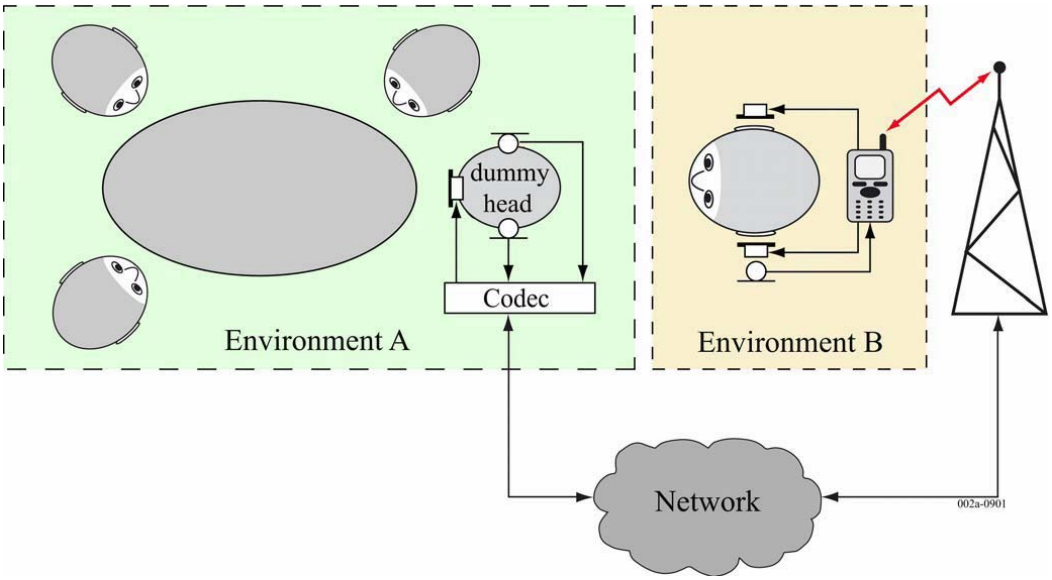

Bild 7: Konzept der binauralen Telefonie

Literatur

[1] Vary, P., Martin, R.: Digital Speech Transmission – Enhancement, Coding and Error Concealment, John Wiley&Sons LTD, 2006

[2] ETSI Rec. GSM 06.90: Digital Cellular Telecommunications System; Adaptive Multi-Rate (AMR) Speech Transcoding, 1998

[3] 3GPP TS 26.190: AMR Wideband Speech Codec; Transcoding Functions

[4] Jax, Peter; Vary, Peter: Bandwidth Extension of Speech Signals: A Catalyst for the Introduction of Wideband Speech Coding? *IEEE Communicat. Magazine*, Mai 2006

[5] Geiser, B. und Vary, P.: Backwards Compatible Wideband Telephony in Mobile Networks: CELP Watermarking and Bandwidth Extension, in *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)* Band IV, (Honolulu, Hawai'i, USA), S. 533–536, Apr. 2007.

[6] Adrat, M., Vary, P.: Iterative Source-Channel Decoding: Improved System Design Using EXIT Charts, *EURASIP Journal on Applied Signal Processing*, Band 2005, S. 928–941, Mai 2005. – Special Issue on Turbo Processing –

[7] 3GPP TS 26.290: Extended Adaptive Multi-Rate Wideband (AMR-WB+) C; Transcoding Functions, 2005

[8] Krüger, H. und Vary, P.: A New Approach for Low-Delay Joint-Stereo Coding, in *ITG-Fachtagung Sprachkommunikation* (Aachen, Germany), Okt. 2008.