

Der Speex-Audiocodec



Agenda

- Einführung / Einordnung des Codec
- Merkmale
- Signalformkodierer vs. parametrischer Kodierer
- Der CELP-Algorithmus
 - Herkunft: Das Modell der Spracherzeugung
 - Dekodierer-Aufbau
 - Synthese-Parameter
 - Kodierungsverfahren
- Speex-Besonderheiten
- Anwendungsgebiete

Einführung

- Der Speex-Codec ist ein Open-Source-Audiocodec speziell für die Sprachübertragung
- Die Lizenz des Codec erlaubt eine Nutzung im kommerziellen Bereich ohne „viralen Effekt“
- Der Codec basiert auf dem CELP-Verfahren und modelliert die menschliche Spracherzeugung
 - Daher hohe Kompressionsraten bei Sprache, aber deutliche Qualitätsverluste bei Musik o.ä.
- Speex wurde primär für VoIP entwickelt, nicht für den Mobilfunk

Merkmale des Codec

- Datenraten zwischen 2 und 44 kBit/s
- Dynamischer Bitratenwechsel möglich
- Abtastraten zwischen 8 und 48 kHz
- Algorithmische Verzögerung von 30-34 ms
- Integrierte Sprechpausenerkennung
- Wählbare Codierungs-Komplexitätsstufen
- Paketverlust-Verschleierung
- Echo-Unterdrückung

Warum Stimme nicht mit mp3 übertragen?

- Natürlich wäre es möglich, reine Sprachsignale (z.B. Telefongespräch) per mp3/Vorbis/wma/... zu codieren und zu übertragen
- Aber: bei Sprachübertragung steht oft ein geringer Bandbreitenbedarf bei ausreichender Sprachverständlichkeit im Vordergrund
- Vergleich: mp3 \sim 128kbit/s, GSM 13 kbit/s
- eine Modellierung der Sprachsignal-Erzeugung ermöglicht höhere Kompressionsraten bei Erhaltung der Verständlichkeit

Signalformkodierer vs. parametrischer Kodierung

Signalformkodierer

Codierung ohne Wissen um die Art der Signalerzeugung

Bitrate typisch 64-640 kbit/s

Bei ausreichender Bitrate originalgetreue Wiedergabe

parametrischer Kodierer

Codierer nimmt an, dass Signal auf eine bestimmte Weise erzeugt wurde, und nutzt ein Modell dieser Erzeugung zur (De-)Codierung

Bitrate bei unter 2.4 kbit/s

Sprache zwar erkennbar, aber keine originalgetreue Wiedergabe

Modell der Spracherzeugung

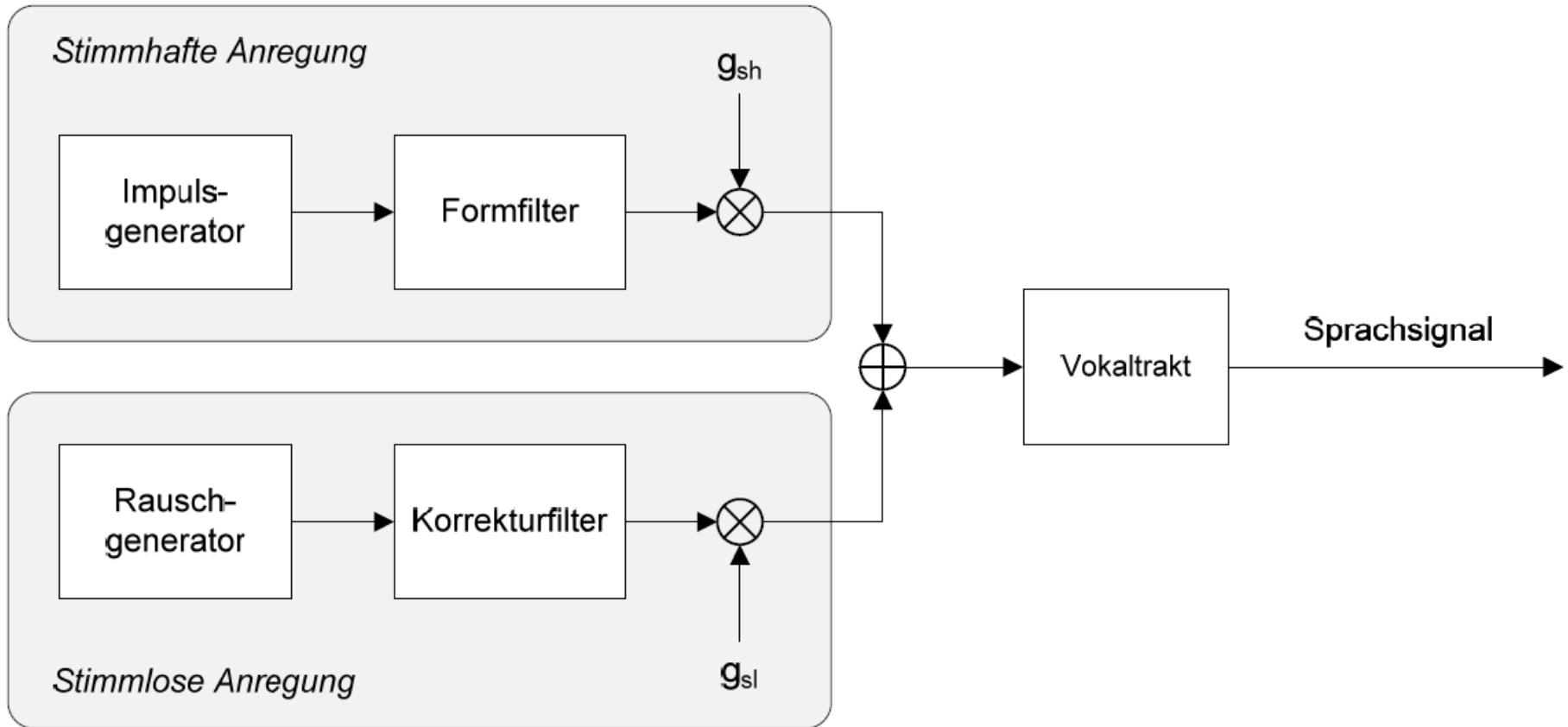
- Die Spracherzeugung im menschlichen Sprechtrakt kann in einem einfachen Modell angenähert werden
- Das Modell besteht aus 2 Teilen: Die Anregung und die Signalformung

Modell der Spracherzeugung

- Es gibt 3 verschiedene Arten der Anregung:
 - **Stimmhaft:** Die Stimmbänder werden durch den Luftstrom der Lunge zum Schwingen gebracht. Amplitude/Periode sind variabel.
 - **Stimmlos:** Der Luftstrom erzeugt ein Rauschen im Mund- und Rachenraum.
 - **Transient:** Der Sprechtrakt wird kurzzeitig verschlossen, wodurch Druck aufgebaut wird, der anschließend entweicht.
- Mischformen sind möglich

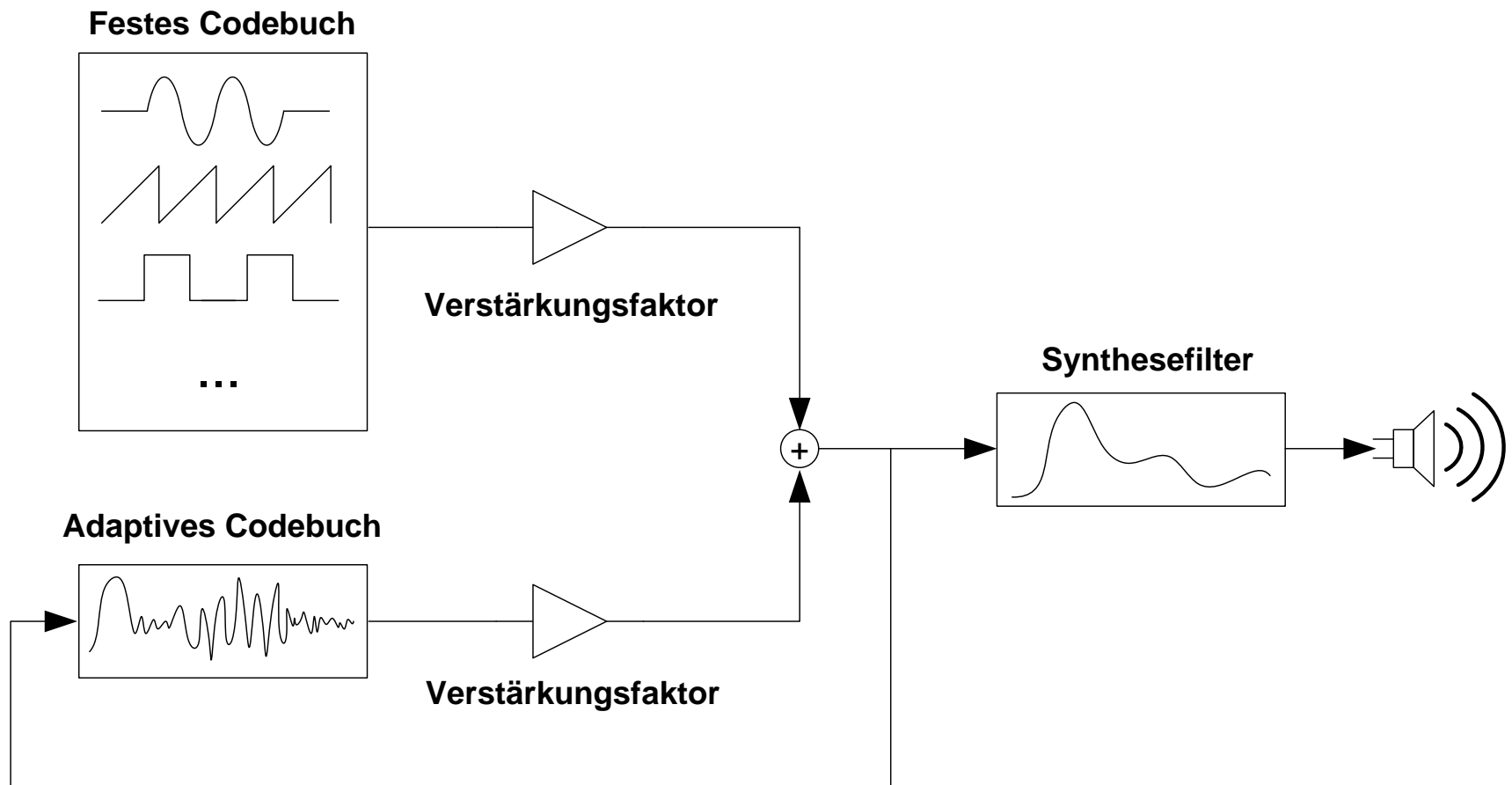
Modell der Spracherzeugung

- Der Hohlraum des Sprechtraktes dient als Resonanzkörper. Durch die variable Geometrie dieses Hohlraums kann das Anregungssignal variabel gefiltert werden.
- Dieser Hohlraum kann vereinfacht als Rohr mit veränderlichem Querschnitt modelliert werden.



Der CELP-Algorithmus

- Grundlagen stammen aus dem Jahr 1985
- Ausgangspunkt für die meisten heute genutzten Sprachcodecs
- Das CELP-Verfahren basiert auf der Synthese eines Audiosignals mittels übertragener Parameter, enthält aber auch Elemente der Signalformkodierung
 - CELP ist daher ein „hybrides Verfahren“



Zu übertragende Parameter

- Verzögerungswert des adaptiven Codebuchs
- Indizes ins feste Codebuch
- Verstärkungsfaktoren
- Koeffizienten des Linearprädiktionsfilters

- Die Übertragung findet getrennt für jeden Subframe (5 msec) statt
 - Ausnahme: Die Koeffizienten für den Linearprädiktionsfilter gelten für einen kompletten Frame (20 msec)

Kodierungsverfahren

- Die CELP-Kodierung basiert auf dem Prinzip der Analyse durch Synthese
- In einer geschlossenen Decoder-Schleife wird ein dekodiertes Signal so lange durch Veränderung der Parameter optimiert, bis es dem zu analysierenden Signal bestmöglich entspricht

Kodierungsverfahren

- Theoretisch sind per Brute-Force optimale Parameter für ein gegebenes Signal ermittelbar -> zu zeitaufwendig
- Stattdessen wird die Analyse in kleine Teile aufgeteilt und eine Gewichtungsfunktion für den Kodierungsfehler angewandt
- Diese Gewichtung sorgt dafür, dass Kodierungsfehler hauptsächlich in wenig bis gar nicht hörbaren Frequenzbereichen auftreten

Speex-Besonderheiten

- **Voice Activity Detection**
 - Erkennt Sprechpausen und encodiert in diesen Pausen lediglich Hintergrundrauschen mit sehr geringer Bitrate (Comfort Noise Generation)
- **Variable Bit-Rate**
 - Durch dynamisches Anpassen der Encoding-Bitrate an das Signal wird Qualität und Bandbreitenbedarf optimiert

Speex-Besonderheiten

- Integrierter Präprozessor
 - Rauschunterdrückung
 - Automatische Signalstärkeregelung
 - Echo-Unterdrückung
- Adaptiver Jitter-Buffer
 - Der Jitter-Buffer korrigiert falsche Paketreihenfolgen und passt seine Größe dabei dynamisch an die minimalen Erfordernisse an.

Anwendungsgebiete

- VoIP-Konferenzapplikationen
 - TeamSpeak
 - Mumble
- VoIP-Support in Computerspielen
 - Counter-Strike
 - Xbox Live
- Diverse IP-Telefonie-Applikationen
- US-Armee im „Land Warrior“-Projekt

Links

- The Speex Codec Manual, Jean-Marc Valin
 - <http://www.speex.org/docs/manual/speex-manual.pdf>
- Implementierung des AMR-Sprachcodecs in Matlab/Simulink, Bachelorarbeit, Sebastian Knop
- Wikipedia: Code Excited Linear Prediction
 - http://en.wikipedia.org/wiki/Code_Excited_Linear_Prediction
- http://www-mobile.ecs.soton.ac.uk/speech_codecs/index.html