

Digitale Sprachübertragung mit niedrigen Bitraten mittels CODEC2 (1)

Von SSB zu den Verfahren der Neuzeit

Prof. Dr. Michael Hartje, DK5HH

In dieser Beitragsserie wird über Verfahren berichtet, mit denen es möglich ist, digitale Sprachübertragung über sehr kleine Bandbreiten, wie sie bei Kurzwellen zugelassen sind, vorzunehmen. Die Serie nimmt sich den grundlegenden Anforderungen an, aber auch der Spracherzeugung und ihrer effektiven Codierung.

Die Sprachübertragung im Amateurfunk auf Kurzwellen hat sich in den vergangenen Jahrzehnten mit SSB (Single Side Band) praktisch nicht verändert. Viele technologische Entwicklungen haben dazu beigetragen, den Empfang störungsfreier und besser zu machen; SSB selbst wurde bisher nie infrage gestellt. Dem gegenüber wurden im VHF- und UHF-Bereich eine Vielzahl von neuen digitalen Sprachübertragungsverfahren in den Amateurfunk eingeführt und haben sich teilweise recht erfolgreich entwickelt. APCO25, D-Star, DMR konnten sich in den vergangenen Jahren etablieren und werden teilweise intensiv genutzt. Die bisherige Technik der SSB- und FM-Übertragung haben sie jedoch auch auf VHF/UHF nicht ablösen, sondern nur ergänzen können. Auch im VHF- und im UHF-Bereich ist SSB für die Sprachübertragung unter schwierigen Bedingungen in der Nähe der Rauschgrenze bisher nicht sinnvoll zu ersetzen.

Die digitalen Übertragungsverfahren werden teilweise aus dem kommerziellen Bereich übernommen und größtenteils erst für den Relaisbetrieb, also im Wettbewerb zur FM-Übertragung über Relais, eingesetzt. Dabei werden teilweise auch zentral gesteuerte Zeittakte für die Koordination der Übertragung zu den mobilen Einheiten genutzt.

Digitale Übertragungsverfahren mit ihren geringeren Bandbreiten des Funkkanals sind vor dem Hintergrund der im Kurzwellenbereich zur Verfügung stehenden Bandbreiten der „Funkkanäle“ und mit

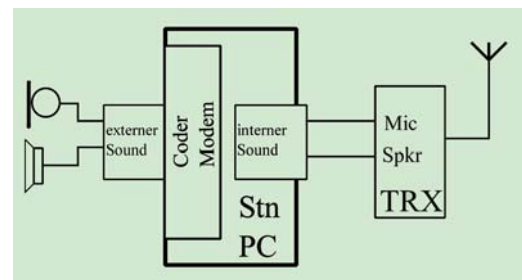
den Anforderungen des Kurzwellenfunks nicht besonders gut vereinbar. Verschiedene Versuche in der Vergangenheit, digitale Sprachübertragung mit sehr niedrigen Bitraten für Kurzwellen zu entwickeln und einzuführen, sind häufig an dem hohen Gerätepreis und der damit einhergehenden geringen Nutzerzahl gescheitert. Dem gegenüber wurde im Wesentlichen von David Rowe, VK5DGR, unter Mitwirkung von international engagierten Funkamateuren in der Vergangenheit ein lizenzfreies Verfahren entwickelt, mit dem es möglich ist, bei sehr niedrigen Bitraten mit ca. 1...2 kBit/s, eine Sprachübertragung mit ausreichend hoher Qualität zu ermöglichen [10]. Im Unterschied zu allen anderen bisher bekannten Verfahren ist es hier ohne die Entrichtung von Lizenzgebühren möglich, die vorhandenen Quellcodes von der Projektseite zu laden, zu übersetzen und lizenzfrei zu betreiben. Dazu ist nur ein Rechner mit zwei Soundkarten erforderlich. Embedded Systems sind inzwischen so leistungsfähig, dass sie ebenfalls die Aufgabe erfüllen können.

Anforderungen der Sprachübertragung

Sprache muss ohne Unterbrechungen möglichst kontinuierlich und mit wenig Verzögerung übertragen werden. Der Frequenzbereich muss ausreichend sein, die Phasenverzerrungen sollen klein sein, sodass die Vokale und Konsonanten voneinander zu unterscheiden sind. Für den Kommunikationsstandard hat sich die Bandbreite von ca. 300 Hz bis

etwa 3,5 kHz aus dem Telefonstandard hieraus entwickelt.

Die analoge Sprachübertragung erfordert infolge der relativ hohen Bandbreite auf der einen Seite einen höheren Signal/Rauschabstand, als Telegrafie oder Textbetriebsarten. Auf der anderen Seite erfordert der Sender relativ hohe Linearität, um die Verzerrungsprodukte niedrig zu halten und dabei eine ausreichende Sprachverständlichkeit zu erzielen.



Das menschliche Gehör und die dazugehörigen Gehirnregionen sind enorm leistungsfähig. Wir haben genetisch bedingt und durch eine intensive Lernphase beginnend mit dem Babyalter die Fähigkeit entwickelt selbst

- unter schwierigen akustischen Bedingungen, wenn Störgeräusche einen Teil der für die Spracherkennung notwendigen Frequenzanteile maskieren (überdecken),
- bei geringem Signal/Rauschabstand (Signal/Störabstand), wenn der Schall „extrem leise“, in der Nähe des „Eigenrauschens durch den Blutkreislauf“ ist,
- mit unterschiedlichen Frequenzlagen der Stimme, Männer-, Frauen- und manchmal sehr hohe und schrille Kinderstimmen,
- bei physisch sehr unterschiedlichen „Stimmbildungsapparaten“, die zur Folge haben, dass die „Klangfarbe“ der Stimme sich sehr stark verändert hat – z.B. nach übermäßigem Alkoholenuss oder starker Dauerbelastung durch lautes dauerhaftes Schreien,
- bei stimmloser Sprache – z.B. beim Flüstern

Bild 1: Stationskonzept Funkgerät mit Erweiterung

Den Autor erreichen Sie unter:
Prof. Dr. Michael Hartje,
DK5HH
Nettelbeckstr. 28
28201 Bremen
dk5hh@darcd.de

eine eindeutige Einordnung der Stimme und ein „Zuhören“ und Verstehen der gesprochenen Worte und Sätze zu ermöglichen.

Wenn es um das „Verstehen“ geht, haben wir gelernt, sowohl unterschiedliche Charakteristiken des Übertragungskanals auszugleichen, als auch Störungen des Übertragungskanals auszublenken.

Um das Verstehen von Sprache nachzuvollziehen, gibt es den analytischen Ansatz, nach den kleinsten Einheiten der Sprache zu suchen. Sprache lässt sich in kleine Phoneme zerlegen. Also Laute, aus denen die Elemente der Sprache zusammengesetzt werden. Das Gehör ist darauf trainiert, sie zu identifizieren und zu den höher gruppierenden Sprachelemente (Silben, Worte, Sätze) zusammenzusetzen.

Das Besondere des „Hörens“ ist es nun im weiteren, dass wir in der Lage sind, auch beim Fehlen einzelner Phoneme Worte zu erkennen. Selbst dann, wenn einzelne Worte nicht mehr erkannt werden, versucht das Gehirn den aus den Worten gebildeten Satz als Ganzes zu verstehen. Dabei nimmt es die Möglichkeit einer Sinnbildung zu Hilfe – was will mir der Sprecher mitteilen? Dieses Verfahren klappt sehr gut – unser Gehirn hat das sehr lange trainiert.

Fortentwicklung der Sprachübertragung per Funk

Die Sprachübertragung auf unterstem Niveau – also lediglich die Wort- und Satzverständlichkeit – ist im Amateurfunk nur bei Contestbetrieb akzeptiert. Will man eine entspannte Unterhaltung führen, ist diese Qualität nicht ausreichend. Schließlich soll nicht nur der Sprecher, sondern auch dessen emotionale Regung – die Stimmung – bei dieser Stimmübertragung mit übertragen werden. Schaut man aufmerksam auf den VHF- und UHF-Bändern herum, so wird sogar deutlich, das manch ein OM/YL nach der ersten Euphorie sich von der digitalen Sprachübertragung, häufig aus Qualitätsgründen, wieder abgewandt hat. Befragt man diese, führen sie ins Feld, dass die Qualität einer guten FM-Übertragung über Relais mit einem sehr ausgewogenen NF-Frequenzspektrum bei ausreichend hoher Feldstärke meist wesentlich angenehmer/unangestrengter ist, als die digitale Sprachübertragung bei D-Star oder DMR, selbst wenn diese noch so viele Vorteile aufzuweisen scheinen und die

lästigen „Rauscheinbrüche“ der mobilen Gesprächspartner hier fehlen.

Für die Kurzwelle haben sich die im VHF- und im UHF-Bereich etablierten digitalen Verfahren noch nicht übertragen lassen. Zwar gibt es eine Reihe von Versuchen dazu, die schon erprobten digitalen Sprachübertragungsverfahren für den KW-Amateurfunk anzupassen. Über den Versuchsbetrieb sind diese Verfahren bisher nicht hinausgekommen. Es gibt wenige Nutzer, die Verfahren sind alle sehr unterschiedlich, manchmal sind erhebliche Aufwendungen erforderlich und schließlich ist wegen der hohen Aufwendungen für die Zusatzgeräte bisher eine kritische Menge von Nutzern nicht erreicht worden. Dies könnte mit dem im Folgenden näher erläuterten Verfahren deutlich anders sein.

Viele Nutzer durch geringen Aufwand

Wegen der fehlenden Lizenzkosten und der wenig aufwändigen, zusätzlichen Geräte besteht die Möglichkeit einer deutlich größeren Verbreitung und damit eine kritische Menge von Nutzern zu übersteigen, die den Einsatz für alle weiteren Funkamateure auf Kurzwelle attraktiv machen.

Möglicherweise steht der KW-Amateurfunk derzeit an einer Schwelle. Schaut man zurück auf die Zeit, in der der Übergang von Amplitudenmodulation zu SSB war, so werden sicherlich die älteren Funkamateure eine sehr emotionalgeführte Diskussion über die Qualität der „Nasenklammermodulation“ erinnern.

Und auch heute ist die Diskussion über die Qualität der Sprachübertragung mit SSB nicht verstummt, wenn man einmal anschaut, wie lange individuell an der Kombination Stimme und Mikrofon im Einzelfall experimentiert wird, um einen möglichst charakteristischen Klang, der die Eigenschaften des Sprechers wiedergeben soll, zu erzeugen. Die Zahl der Artikel in den Amateurfunkpublikationen, die sich mit der Modulation bei FM oder SSB befassen, ist erschlagend.

Nicht nur der relativ geringe Aufwand für das weiter unten beschriebene neue Verfahren, sondern die vermutlich in absehbarer Zeit verfügbaren sehr einfachen Zusatzgeräte könnten diesen angedeuteten Umbruch im Kurzwellen-Sprechfunk herbeiführen oder zumindest stark begünstigen.

Konzept für die Erweiterung der vorhandenen Funkgeräte

Bei neuer Technik sollte die Schwelle im Amateurfunk sowohl hinsichtlich der apparativen und finanziellen als auch der organisatorischen Aufwendungen möglichst gering gehalten werden. Nur dann gelingt es, möglichst viele zum Ausprobieren zu gewinnen. Der einfachste Fall ist sicherlich, wenn ein bestehendes Funkgerät durch eine Zusatzeinrichtung für die weitere Betriebsart nachgerüstet werden kann. In der Vergangenheit haben sich viele digitale „Betriebsarten“ (Modulationsverfahren), wie PSK31, RTTY, Olivia, MT63, WSPR und auch das analoge SSTV mithilfe eines Stations-PCs und der eingebauten Soundkarte realisieren lassen. Dabei war meist die Tastatur und der Bildschirm Quelle und Senke der Informationen.

Für eine digitale Sprachübertragung müsste dieses Konzept um eine eigene Soundkarte erweitert werden, die die Schnittstelle zu Quelle und Senke für die Sprachinformationen (Mikrofon, Lautsprecher) realisieren kann (**Bild 1**). Externe USB-Soundkarten kosten teilweise nur 1,50 €. Sie erfüllen voll und ganz die gewünschten Anforderungen hinsichtlich Linearität und Dynamikumfang für den Sprachbetrieb. Damit werden Mikrofon und Lautsprecher nun nicht mehr unmittelbar an den Transceiver, sondern an die externe Soundkarte des Stations-PCs angeschlossen. Das Programm „Coder Modem“, im Rechner ausgeführt, kann sowohl die digitale Codierung, als auch die Modemfunktionen bereitstellen. Die interne Soundkarte dient dazu, diese aufbereiteten Tonsignale an den Transceiver (Mikrofon- und Lautsprecheranschluss) zu übertragen.

Der vorhandene SSB-Sender (Trx) setzt die digitalen Signale (Tonsignale) in den Kurzwellenbereich um und gibt diese verstärkt an die Antenne.

Neuere Ansätze zeigen Wege auf, mit denen statt eines Stations-PCs ein externes Zusatzgerät eingesetzt werden kann, welches die gleichen Funktionen übernimmt, aber sowohl kleiner in den räumlichen Abmessungen ist, als auch deutlich weniger Leistung verbraucht. Derzeit wird in einer internationalen Zusammenarbeit von Funkamateuren daran gearbeitet, ein Speaker-Mikrofon mit dem entsprechenden Zusatzgerät für Codierung und Modembetrieb zu entwickeln.

(Fortsetzung auf S. 9)

Digitale Sprachübertragung mit niedrigen Bitraten mittels CODEC2 (2)

Anforderungen der digitalen Sprachübertragung für KW

Prof. Dr. Michael Hartje, DK5HH

Digitale Sprachübertragung knüpft an das Modell der analogen Sprachübertragung an. Daher ist unmittelbar hinter dem Mikrofon und kurz vor dem Lautsprecher der Übergang von analogen Signalen zu digitalen Signalen mit wertediskreten und zeitdiskreten Darstellungen der „Abtastwerte“.

Die Digitalisierung der Sprache für die ebensolche digitale Übertragung ist ein umfangreiches Wissensgebiet. Es werden Verfahren eingesetzt, die zu einer stark reduzierten, gerade noch tolerierten Qualitätsverminderung führen. Diese wird üblicherweise in standardisierten, subjektiven Tests mit einer großen Zahl von Personen ermittelt. Beurteilungsgrundlage ist zumeist, dass eine Übertragung nur auf den so genannten Kommunikationsstandard abstellt. Dies bedeutet, dass Verzerrungen und Artefakte – künstlich entstehende, im Original nicht vorhandene Klänge – immer noch eine akzeptable Sprachverständlichkeit ermöglichen. Diese wird dann in diesen Tests beurteilt.

Seit Einführung des Telefons ist die eingeschränkte Qualität für diesen Informationsaustausch im Gespräch akzeptiert. Bei der Digitalisierung eines bandbegrenzten Audiosignals werden heute bei der Telefonübertragung üblicherweise eine Bandbreite von 4 kHz und eine nichtlineare Quantisierung eingesetzt. Dabei entstehen meist 64 kBit/s, bei Verzicht auf die nichtlineare Quantisierung sogar 128 kBit/s. Die dafür erforderliche Technik ist relativ einfach. Sie besteht aus einem Analog-Digital-Umsetzer (ADU), einem Modem und einem dazu entsprechenden Empfänger, bestehend aus dem Modem und einem Digital-Analog-Umsetzer.

Kanalkapazität des HF-Kanals

Für die Übertragung eines Datenstromes von 64 oder 128 kBit/s aus der digitalisierten Sprache wird entweder

eine sehr große Bandbreite von vielen 10 kHz oder ein sehr großer Störabstand von mehreren 10 dB – am besten beides – erforderlich. Gerade diese Einschränkungen verbieten eine solche Übertragung auf Kurzwellen.

Die Informations- und Nachrichtentechnik fasst einen für eine Übertragungsaufgabe genutzten Bandbereich als einen „Kanal“ zusammen. Der maximal mögliche Datenstrom, der über diesen Kanal übermittelt werden kann, lässt sich mit der Shannonschen „Kanalkapazität“ [2, 14] für gestörte Kanäle genauer berechnen:

$$C = B \cdot \frac{\ln(1 + \frac{S}{N})}{\ln 2} \quad 1.$$

mit der Näherung für sehr gute Signal-Störabstände der Leistungen, also $S/N \gg 1$

$$C \approx \frac{B}{3} \cdot \frac{S}{N} \Big|_{dB} \quad 2.$$

Die Gleichung 2 lässt sich nur bei Fällen mit großen Signal-/Störabständen von mehr als 10 dB als Näherung verwenden, sodass im Folgenden zunächst mit der Gleichung 1 gerechnet werden soll. Auf KW ist heute mit SSB eine Übertragung mit wenigen dB Signal-/Störabstand bei einer Bandbreite von ca. 2,4 kHz für ein gut trainiertes Gehör eines guten Op möglich. In den USA ist die maximal zulässige Bandbreite auf 2,7 kHz für den SSB-Betrieb begrenzt, in Europa ist das m.W. nicht so streng geregelt, sodass Versuche mit größerer Bandbreite grundsätzlich möglich sind. Damit wird die Kanalkapazität bei konstanten Verhältnissen und dem

Zur Person



Prof. Dr. Michael Hartje, DK5HH
 Anschrift:
 Nettelbeckstr. 28
 28201 Bremen
 dk5hh@dar.c.de

geringen Leistungsverhältnis des Nutzsignals zu den Störungen mit $S/N = 2 = 3$ dB mit $B = 2,7$ kHz zu

$$C_{2,7 \text{ kHz}} \leq 2700 \cdot \frac{\ln(1 + 2)}{\ln 2} = 4280 \text{ Bit/s} \quad 3.$$

Diese theoretische Grenze wird jedoch in der Praxis nicht erreichbar. Auch die digitalisierte Sprachübertragung darf derzeit höchstens die zugelassene Bandbreite nutzen. Bei dem in der Kurzwellen manchmal geringen Signal-Störabstand und mit den ständig sich verändernden Eigenschaften der Übertragung lassen sich in einem solchen „Kanal“ (gemeint ist die nutzbare Bandbreite) damit maximal 1...3 kBit/s einigermaßen sicher übertragen.

Bei einer geringeren Bandbreite von z.B. $B = 1,2$ kHz, wie sie für die digitale Sprachübertragung mit FreeDV [4] benötigt wird, entsteht die theoretische Kanalkapazität von

$$C_{1,2 \text{ kHz}} \leq 1200 \cdot \frac{\ln(1 + 2)}{\ln 2} = 1902 \text{ Bit/s} \quad 4.$$

Es ist zu erkennen, dass die von FreeDV genutzte Übertragungsrate von 1600 Bit/s diesem theoretischen Wert schon recht nahe kommen wird. Bei einem konstanten Signal-/Störabstand von etwa 10 dB erhöht sich die theoretisch erreichbare Kanalkapazität dann auf

$$C_{1,2 \text{ kHz}, 10 \text{ dB}} \leq 1200 \cdot \frac{\ln(1 + 10 \cdot 10^{\frac{10}{10}})}{\ln 2} = 5271 \text{ Bit/s} \quad 5.$$

Wegen der starken Schwankungen mit frequenzselektivem Fading ist aber auch dieser Wert nicht zu erreichen. Die Kanalkapazität von Funkkanälen auch

mit Fading ist z.B. in [13] recht gut beschrieben und noch einmal in [2] zusammengefasst. Wenn das zeitlich wiederholt auftretende Fading durch eine Verzögerung in der Signalverarbeitung und einen Puffer ausgeglichen werden kann, lassen sich die aus der Theorie abgeleiteten Kanalkapazitäten besser erreichen.

Für die Echtzeitanwendung mit digitalisierter Sprache sind die theoretisch abgeleiteten Werte bei Fading vor allem deswegen nicht erreichbar, weil die Sprache ohne Verzögerung übertragen werden soll, und um dann ein schnelles Wechselsprechen zu ermöglichen. Grob geschätzt und ohne weiteren Nachweis kann man die „erreichbare Kanalkapazität“ aus der Kurzwellenpraxis heraus ableiten. In den kommenden Beiträgen der Serie ist aber noch an den praktischen Beispielen mit Fading nachzuvollziehen, dass diese praktisch „erreichbare Kanalkapazität“ geschätzt eher bei

$$C_{\text{praktisch, 1,2 kHz, 10 dB}} \approx \frac{C_{1,2 \text{ kHz, 10 dB}}}{3} \approx 1757 \text{ Bit/s} \quad \mathbf{6.}$$

liegt. Für die digitale Sprachübertragung streben wir daher z.B. 1600 Bit/s bei einem Signal-/Störabstand von im Mittel 10 dB an. Damit ist auch bei kleinem Fading dann schon einen recht sicheren und wenig gestörter Betrieb zu erreichen. Verwendet man die Näherungsgleichung 2, so ergibt sich

$$C_{1,2 \text{ kHz, 10 dB, genähert}} \approx \frac{1200 \cdot 10}{3} = 4000 \text{ Bit/s} \quad \mathbf{7.}$$

als Maximalwert, der bei dem unvermeidlichen Fading der Kurzwellenübertragung wieder auf einen deutlich kleineren Wert absinkt und zu der praktisch „erreichbaren Kanalkapazität“ von

$$C_{\text{praktisch, 1,2 kHz, 10 dB, genähert}} \approx \frac{C_{1,2 \text{ kHz, 10 dB, genähert}}}{3} \approx 1333 \text{ Bit/s} \quad \mathbf{8.}$$

führen würde. Die Näherungsgleichung 2 führt erst mit Signal-/Störabständen von 20 dB zu einem ähnlichen Wert, wie die exakte Berechnung. Die Berechnungen und der Vergleich zeigen, dass wir etwa mit diesen Kapazitäten zwischen 1330 und 2000 Bit/s bei einer Kurzwellenübertragung sinnvoller Weise rechnen sollten.

Das bedeutet für den Kurzwellenfunk, dass wir hier mit sehr kleinen Übertragungsraten zufrieden sein müssen. Wir wollen uns daher der Frage zuwenden, wie wir von den anfänglich 128 kBit/s der digitalisierten Sprache auf diese niedrigeren Übertragungsraten kommen können.

Digitalisierte Sprache als Echtzeitanwendung

Durch eine so genannte Quellencodierung werden die anfänglich bei der Digitalisierung mit dem Analog-Digital-

Umsetzer (ADU) erzeugten 128 kBit/s auf deutlich niedrigere Bitraten transformiert/codiert. Diese Codierung muss in der Größenordnung von Faktor 100 die Datenrate reduzieren können, damit im Kurzwellenbetrieb eine digitale Modulation mit zulässiger Bandbreite die Information vollständig übertragen kann. So entsteht ein binärer Datenstrom von wenigen Bytes pro Sekunde, der alle Sprachinformationen enthält. Dieser muss fehlerlos übertragen werden. Als Besonderheit der digitalen Sprachübertragung gegenüber der bislang im Amateurfunk üblichen binären Datenübertragung ist hier zu nennen, dass diese digitale Sprachübertragung in Echtzeit erfolgen muss. „Realtime“ wird allgemein so definiert, dass die zu übertragenden Daten nur während eines bestimmten Zeitraums Gültigkeit haben. Danach müssen sie verfallen.

Die Netzprotokolle, die bisher im Amateurfunk eingesetzt wurden, haben die Eigenschaft, dass sie als verbindungsbehaftete Protokolle, wie beim Netzwerkprotokoll TCP, in kleine Pakete zerlegt übertragen werden. Diese können durch eine Checksumme des Datenpaketes auf Richtigkeit geprüft werden. Im Fehlerfall wird auf dem Rückkanal der erkannte Fehler gemeldet und das Datenpaket kann wiederholt werden. Die Nutzer von Packet Radio kennen das Wiederholen der Pakete als eine typische Eigenschaft von Packet Radio un-

Literatur

Es wurden überwiegend frei zugängliche Quellen verwendet:

- [1] Bonne, K. Specification of AOR/ADR9000 modem., http://villazeebries.krbonne.net/hamstuff/?page_id=86, 3/13
- [2] Martin Bossert, Klaus-Dirk Kammeyer: „Nachrichtenübertragung“, 5. Aufl., Wiesbaden, Vieweg + Teubner, 2011 (Studium).
- [3] Mäusl, R.; Göbel, J.: „Analoge und digitale Modulationsverfahren: Basisband und Trägermodulation“, Heidelberg: Hüthig, 2002.
- [4] FreeDV HF Digital Voice for Radio Amateurs, <http://freedv.org>. Webabruf: 3/13
- [5] SoX - Sound eXchange Homepage, <http://sox.sourceforge.net>, Webabruf: 3/13
- [6] Die Stimme - Anatomie & Physiologie, www.vocalis-projekt.de/html/stimme/exkurs.htm#kehlkopf, Webabruf: 8/13, Kehlkopfbild nach v. Essen
- [7] TWELP 1200 bps Vocoder – Datasheet, <http://twelp.pro/DSPINI.Datasheet.TWELP.1200bps.Vocoder.pdf>, Webabruf: 7/13
- [8] WinDRM 1.3 Dokumentation and Usermanual, http://n1su.com/windrm/Windrm_docs_Release1r3.pdf, Webabruf: 4/13
- [9] Pompe, Bernd: „Die Spracherzeugung beim Menschen“, <http://www2.physik.uni-greifswald.de/~pompe/SCRIPTS/habil-kolloquium-pompe.pdf>, Webabruf: 6/05
- [10] Rowe, David, VK5DGR: CODEC2, www.rowetel.com/blog/?page_id=452
- [11] Rudolph, Dietmar: „Vielträgermodulation“, www.diru-beze.de/funksysteme/skripte/DiFuSy/DiFuSy_OFDM_WS0405.pdf, Webabruf: 3/10
- [12] Smith, Julius O.: „Physical audio signal processing for virtual musical and audio effects - Dudley's Vocoder“, https://ccrma.stanford.edu/~jos/pasp/Dudley_s_Vocoder.html, Webabruf: 8/13
- [13] Tse, D.; Viswanath, P.: „Capacity of wireless channels“, Cambridge University Press, 2005, www.eecs.berkeley.edu/~dtse/Chapters_PDF/Fundamentals_Wireless_Communication_chapter5.pdf
- [14] Wikipedia: „Kanalkapazität des AWGN“ (additives weißes gaußsches Rauschen), <http://de.wikipedia.org/wiki/Kanalkapazit%C3%A4t>. Webabruf: 3/13
- [15] Wikipedia: „Speex“, <http://de.wikipedia.org/wiki/Speex>, Webabruf: 3/10
- [16] Wikipedia: „Golay-Code“, <http://de.wikipedia.org/wiki/Golay-Code>, Webabruf: 3/13
- [17] Wikipedia: „Opus (Audioformat)“, [http://de.wikipedia.org/wiki/Opus_\(Audioformat\)](http://de.wikipedia.org/wiki/Opus_(Audioformat)), Webabruf: 4/13

ter schwierigen Bedingungen. Wenn wir damit Sprache übertragen wollten, hätte das erhebliche Auswirkungen: Diese gesicherte Nachrichtenübertragung hätte bei schwierigen Übermittlungsverhältnissen zur Folge, dass einzelne Teile erneut oder gar mehrfach wiederholt übertragen werden müssten. Da es sich aber um eine Echtzeitübertragung mit einer sehr knapp ausgelegten kleinen Datenrate handelt, fehlt die Zeit für diese Wiederholung und auf der Empfangsseite läuft buchstäblich die Zeit davon. Die damit digital übertragene Sprache würde plötzlich „abgehackt“, wenn ein Datenpaket nicht fehlerfrei übertragen wird – der Audiodatenstrom mit den kontinuierlich zu liefernden Abtastwerten für den Digital-Analog-Umsetzer (DAU) am Lautsprecher wird unterbrochen. Unser Hörzentrum im Gehirn ist in der Lage, teilweise „abgehackte“ Sprache

zusammenzusetzen, und die fehlenden Teile zu ergänzen. Wir überhören diesen zeitlichen Verzögerungsfehler, weil wir mit der Erkennung von Worten den Sinn zusammenstellen können und dann den Sprecher trotz der Fehler noch verstehen. Für die digitale Übertragung von Sprache werden wir diese Eigenschaft später noch benutzen. Die üblichen Übermittlungsverfahren mit Netzwerkprotokollen, die mit Datenpaketen und deren Absicherung durch Checksummen arbeiten, sind damit nicht unser Favorit, wir werden sie hier nicht verwenden wollen. Ähnlich dem Netzwerkprotokoll UDP werden die Audiopakete von Abtastwerten als „Echtzeitpakete“ ohne Kontrollrückfrage übermittelt und dort zum Digital-Analog-Umsetzer (DAU) und zur Lautsprecherausgabe gereicht. Unser Hörzentrum ist in der Lage, zeitlich

kurze, fehlende Teile der Sprache (übersprungene Abschnitte des Zeitverlaufs) besser zu verarbeiten, als der oben erwähnte abgehackte Sprachverlauf mit zwischenzeitlich „stillen Pausen“.

Der erfahrene Kurzwellenamateur weiß, dass er durch gelegentlich auftretende Gewitterstörungen (QRN) oder durch ein gelegentliches auftretendes starkes Knacken in dieser Zeit nichts hören kann, aber letztlich das QSO auch unter diesen erschwerten Bedingungen zu führen ist. Die digitale Sprachübertragung auf Kurzwelle stellt also einige zusätzliche Anforderungen, die bei der binären Datenübertragung von Text oder Standbildern fehlen. Diese Verfahren waren in der Vergangenheit „etwas einfacher“ zu realisieren.

(wird fortgesetzt)



Der dritte Teil befasst sich mit dem Modell der Spracherzeugung und der effizienten Codierung.



Was ist eigentlich . . . JT65-HF?

JT65-HF hat seinen Ursprung im Programm WSJT von Joe Taylor, K1JT, und ist auf die Bedingungen im HF-Bereich modifiziert.

Ein QSO kann noch erfolgreich sein bei sehr guten Bedingungen bei einem S/N-Wert von -25 dB, bei normalen Kanälen und Bedingungen sind es -20 dB. Diese Werte gelten für die Bandbreite eines SSB-Kanals. Der Sender läuft im Dauerstrich!

JT65-HF hat FEC, wird gesendet als AFSK mit 65 Tönen zu je 0,37152 Sekunden Dauer. Das bedeutet, das genau wie bei BPSK infolge einer Übersteuerung (AGC beachten) die gefürchteten „Gartenzäune“ auftreten.

Die guten Werte bei QSOs mit JT65-HF sind allerdings nur durch die Tricks im Übertragungsverfahren möglich. Neben der sehr genauen Zeit ist auch die Anzahl der Bits ein wichtiger Faktor. Zunächst müssen beide Stationen eine sehr genaue PC-Uhr haben, hier spielen Abweichungen von mehr als 1 Sekunde eine bedeutende Rolle. Der Start einer Sequenz beginnt grundsätzlich mit der Sekunde 1 einer Minute, sie endet nach 46,8 Sekunden. Es sind dabei einschließlich FEC

QSO-Beispiel

```
CQ AB1CDE XY12
cq-Ruf von AB1CDE mit Locator

AB1CDE XY1GHK VY73
XY1GHK ruft AB1CDE mit Locator

XY1GHK AB1CDE -15
AB1CDE sendet den Rapport

AB1CDE XY1GHK R-20
XY1GHK bestätigt mit R und Rapport

XY1GHK AB1CDE RRR
AB1CDE bestätigt mit RRR, QSO-Ende

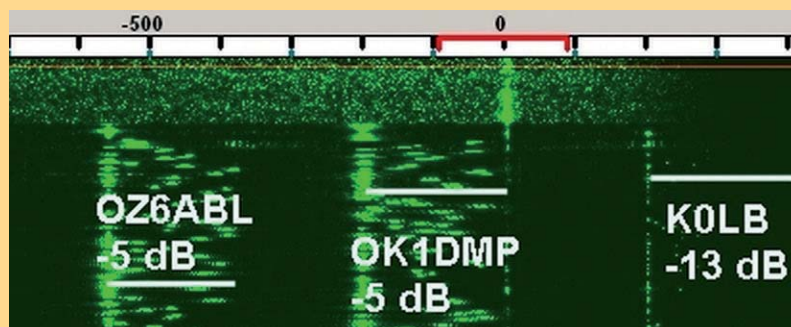
Zusätzlich könnten beispielsweise
noch folgen:

AB1CDE XY1GHK 73
oder
TNX 73 BYE SK (maximal 13 Zeichen
mit Leertaste)
oder
5W BEAM 73 SK
```

378 Bits übertragen worden bei maximal 18 Zeichen. Der Synchronon liegt bei 1275,88 Hz. Es gibt 65 Töne mit je 2,69 Hz Abstand und einer Länge von je 0,37152 Sekunden. Der NF-Bereich umfasst somit 1275,88 Hz bis 1445,46 Hz. Das Programm selbst hat einen Bandpass mit einer Breite von 2 kHz und der Mittenfrequenz 1270 Hz.

Es ist weiterhin festgelegt, was wann übertragen wird. Eine spezielle Sequenz ermöglicht die Übertragung von selbst gewählten 13 Zeichen. Sie darf aber nicht mit RRR, RO oder 73 beginnen. Verwendbar bei JT65-HF sind nur die Zeichen 0...9, A...Z sowie Plus, Minus, Punkt, Schrägstrich und Fragezeichen. Auch das Rufzeichen unterliegt Normen, Sonderrufzeichen oder solche mit Präfix/Suffix haben keine Chance. Der Locator darf nur vierstellig sein. Ein perfektes QSO zeigt der **Kasten**.

Max Perner, DM2AUO



80 m, 0621 UTC, K0LB außerhalb des Bandpasses

Modell der Spracherzeugung für eine effiziente Codierung

Prof. Dr. Michael Hartje, DK5HH

Wir wollen uns zunächst einmal näher ansehen, wie Sprache entsteht und ob sich daraus Modelle ableiten lassen, die für die erforderliche Quellencodierung dienen können. Alles, was wir in dem Modell als Information schon kennen oder was sich irgendwie auf der Rx-Seite herleiten lässt, müssen wir nicht vom Sender zum Empfänger übertragen, da wir diese Informationen als A-priori-Wissen bereits in den Modellen beim Empfänger vorhalten können.

Die Laute der Sprache lassen sich in kleine Einheiten, so genannte Phoneme zerlegen. Im Folgenden sollen die Eigenschaften einzelner ausgewählter Phoneme erläutert werden. Im **Bild 2** ist der Zeitverlauf meines deutsch lang gesprochenen A (oben) und eines E (unten) dargestellt. Die überschwingungsreichen Verläufe haben beide eine Wiederholzeit mit etwa 7 ms. Das entspricht einer Grundfrequenz von etwa 140 Hz. Beide wurden mit etwa der gleichen Tonhöhe gesprochen. Die Amplituden sind geringfügig unterschiedlich, was hier allerdings ohne Bedeutung ist. Das **Bild 3** zeigt den Amplitudengang der beiden Zeitverläufe aus Bild 2. Dabei entspricht 0 dBFS einem Sinus mit der Amplitude 1. Die eingezeichneten Zahlen sind die Ordnungszahlen der Harmonischen $N_{\text{Harmonisch}}$. Mit der Ordnungszahl 1 wird die so genannte Grundschwingung bezeichnet. Die Ordnungszahl der Oberschwingungen berechnet sich nach Gln 9.

$$N_{\text{Oberschwingung}} = N_{\text{Harmonisch}} - 1 \quad (9)$$

Die in dem Bild 3 gezeigten Ordnungszahlen beziehen sich auf den Buchstaben A. Es ist zu erkennen, dass die Grundschwingungen von A und E einige Hertz auseinander liegen. Die Grundschwingung des A beträgt 140 Hz. Der Frequenzunterschied hat zur Folge, dass bei der 10. Oberschwingung von E bereits die 11. Oberschwingung von A erreicht wird (Ordnungszahl 12). Betrachtet man das Spektrum des Buchstabens A, so lässt sich dazu folgendes anmerken:

- Die Grundschwingung liegt bei etwa 140 Hz,
- die Harmonischen 1...6 bilden eine Gruppe mit stärkeren Amplituden,
- besonders stark sind die Amplituden der Harmonischen 1 und 5 mit etwa -25 dBFS,
- die folgende Gruppe liegt bei den Harmonischen 6...9,
- die Amplituden von 3, 6, 9 sind etwa gleich stark mit -35 dBFS,
- die 19. Harmonische des A ist um 20 dB höher als die 11. bis 16. Harmonische.

Der Amplitudengang des Buchstabens E zeigt folgendes:

- Die Amplituden der Harmonischen 1...3 sind relativ stark und bilden praktisch einen Block,
- die Harmonischen 1 und 2 sind fast gleich stark,
- die 3. Harmonische ist gegenüber der Grundschwingung um 8 dB schwächer,

- die Harmonischen 4...15 sind um 20 dB schwächer als die 3. Harmonische,
- die Harmonischen 15...20 sind etwa 20 dB stärker als die Harmonischen 10...13. Sie bilden eine Gruppe,
- die Harmonische 17 hat eine deutlich geringere Amplitude,
- die Harmonischen 15, 16, 18...20 sind wieder mit -40 dBFS deutlich stärker und deutlich gegen über den schwächeren umgebenden Harmonischen zu erkennen.

Wenn wir diese Attribute als charakteristisch ansetzen wollen, ermöglicht ein Vergleich der beiden Spektren auf dieser Grundlage, zu unterscheiden, ob ein A oder ein E gesprochen wurde.

Spektraldarstellung einer Zeitfolge

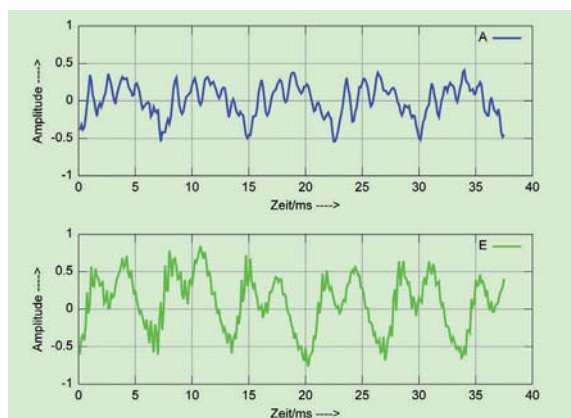
Ein deutlich anderes Verhalten zeigen die stimmlos gesprochenen Buchstaben S und F. Das Spektrum bis 4 kHz ergibt eine etwa konstante Amplitude. Wir sprechen von einem Rauschspektrum. Das **Bild 4** zeigt über eine Sprachaufnahme mit einer Zeit von etwa 5 Sekunden die einzelnen gesprochenen Buchstaben "a e ö ü u o ä f s t". Die vorher in den Bildern 2 und 3 behandelten Buchstaben A und E wurden aus dieser Sprachaufnahme entnommen.

Die einzelnen, in die Tiefe dargestellten Amplitudengänge/Frequenzverläufe wurden jeweils durch eine Fouriertransformation mit 4096 Abtastwerten gewonnen. Das entspricht einer Zeit von 512 ms pro Darstellung. Bei einer Zeit von etwa 2 s wird das U gesprochen (hellblaue Darstellung). Deutlich sichtbar ist hierbei der sehr starke Grundton und die sehr schwachen Oberschwingungen, die bereits ab etwa 1 kHz im „Rauschen“ der Nachbarfrequenzen verschwinden. Es ist bekannt, dass der Buchstabe U, sauber gesprochen, fast nur einen Grundton erzeugt, also einen sinusförmigen Zeitverlauf mit einem sehr geringen Oberschwingungsanteil aufweist. Die vorletzten Frequenzverläufe (bei Zeit = 4 - 4,5 s) in den Farben dunkelblau grün und gelbgrün zeigen die Buchstaben S und F.

Den Autor erreichen Sie unter Prof. Dr. Michael Hartje, DK5HH Nettelbeckstr. 28 28201 Bremen dk5hh@dar.c.de

Die Literaturangaben finden Sie in Ausgabe 2/14, S. 10

Bild 2: Ausschnitt aus dem zeitlichen Amplitudenverlauf eines lang gesprochenen A und eines lang gesprochenen E mit gleicher Tonlage



Deutlich sichtbar ist der sehr gleichmäßige Frequenzverlauf. Dies entspricht einem Rauschspektrum.

Das gesprochene T – ganz vorne bei Sekunde 5, hellblau, – wird als „Teete“ gesprochen. Deutlich sichtbar sind angehobene Amplitudenwerte im Bereich von 2,5...3,5 kHz, die sich aus dem E ableiten und im Vergleich zum zweiten Verlauf bei 0,5 s in diesem Frequenzbereich eine starke Ähnlichkeit aufweisen. Dennoch ist im Vergleich mit dem E die Anhebung deutlich geringer. Das impulsförmig gesprochene T am Anfang wird hierin deutlich. Impulse weisen nahezu gleich verteilte Amplituden auf.

Wenn wir das Bild 4 noch weiter drehen, entsteht ein neue Darstellung, Spektrogramm genannt, wie sie bei der Analyse von Sprache und Tonsignalen bevorzugt verwendet wird. Die Darstellung des Bildes 5 wurde mittels des Programms SoX [5] und der Option „spectrogramm“ aus dem schon vorher verwendeten Sprachbeispiel erzeugt.

Am oberen Rande sind die jeweils gesprochenen Buchstaben zu sehen. Die Zeitachse läuft von links nach rechts über 5,5 s. Die Frequenzachse an der linken und rechten Seite bildet den Frequenzbereich 0...4 kHz ab. Die Amplituden werden als Helligkeitsfarbwerte abgebildet. Die Skala an der rechten Seite des Bildes gibt den Bereich von -10 bis -80 dBFS an. Die Amplitudengänge der schon näher behandelten beiden Buchstaben A und E erleichtern nun die Orientierung und den Vergleich mit den beiden Bildern 3 und 4. Im Bild 3 sehen wir die starken Amplituden bei 140 Hz und die schwächeren niederzahligen Oberschwingungen. Im Bild 5 sind diese über den Zeitraum des gesprochenen A von 300 ms am linken Bildrand als dunkel(blaue) horizontale Linien zu erkennen. Zwischen diesen dunklen Linien weisen die hellen Bereiche auf sehr niedrige Amplituden hin. Dem Zeitraum von 0...300 ms, in denen das A gehalten wurde, folgt zeitlich ein heller senkrechter Bereich – die Pause zum folgenden Buchstaben E. Die Pausen zwischen den anderen Buchstaben sind nicht ganz deutlich zu erkennen.

Die ähnlichen Amplitudenverteilungen für E und Ä werden ebenfalls sichtbar. Sehr schön deutlich ist bei den Buchstaben S und F, die als „esssss“ und „Efff“ gesprochen wurden, dass die ersten 100 ms nach der Pause zum vorherigen Buchstaben die Spektralverteilung etwas anders ist – sie ähnelt beim genauen Betrachten dem Buchstaben E. Das war ja

auch zu erwarten, wenn wir es so sprechen. Der zeitlich hinter dem e folgende Teil von S und F ist dann aber mit relativ gleichmäßiger Amplitudenverteilung (gleicher Farbton ohne dunkle Linien) dargestellt.

Beide Buchstaben unterscheiden sich nur etwas bei etwa 1,5 kHz: Dort hat das F eine etwas gleichmäßigere Amplitudenverteilung, während das S in diesem Bereich einen Einbruch von -25 dB gegenüber den Frequenzbereichen bei 0,5 und 3 kHz aufweist. Diese sehr feinen Unterschiede machen es unserem Hörzentrum im Gehirn möglich, diese Konsonanten S und F zu unterscheiden.

Wir fassen zusammen: Die beiden unterschiedlichen Anregungszustände der Sprache – stimmloses Rauschen und stimmhafte Vokale (Phoneme) – sind die Grundlage für Sprachsynthese und die Modellbildung der synthetischen Spracherzeugung. Eine Unterscheidung der einzelnen Phoneme u.a. ist mittels der spektralen Analyse möglich. Spracherkennungssysteme und Diktiersoftware nutzen diese Erkenntnisse.

Da wir beim Kommunikationsstandard nach der Verständlichkeit schauen, ist eine weitere Betrachtung anhand dieser Bilder möglich. Dabei wollen wir die Verständlichkeit so definieren, dass wir beim Hören möglichst die einzelnen Vokale und Konsonanten gut unterscheiden können.

Bedeutung für die analoge Modulation

Obwohl ein großer Anteil der Sprachenergie im sehr tiefen Frequenzbereich unterhalb 300 Hz liegt, trägt dieser zur Verständlichkeit nur sehr wenig bei. Hingegen weisen die Bereiche zwischen 300 Hz und 4 kHz deutlich geringere Sprachsignalenergie auf, sind aber für die Verständlichkeit – also die Unterscheidung der einzelnen Silben und Phoneme – wesentlich bedeutsamer. Das Bild 3 zeigt, dass die Amplituden der in der Frequenz tief liegenden Amplitudenpitze bei 140 Hz etwa 10 bis 20 dB höher als alle übrigen Anteile sind. Das entspricht einem Unterschied der Leistungen von Faktor 10...100.

Die Sprachenergie der Grundschwingungen bei 140 Hz ist also bei diesen betrachteten Vokalen 10...100 Mal stärker als alle anderen Frequenzanteile der Sprache. Untersuchungen zeigen aber, dass diese auf sehr tiefer Frequenz liegende Sprachenergie für die Verständlichkeit zweitrangig ist, für den wohlklingenden

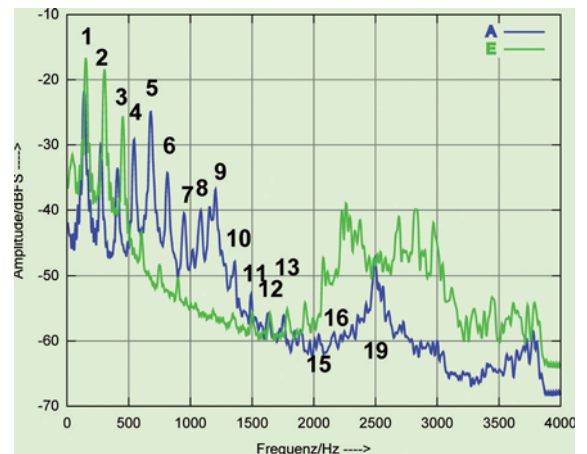


Bild 3: Amplitudengang des lang gesprochenen A und E aus zwei mit benannten Zahlen der Harmonischen zum A. FFT-Länge mit 512 Abtastpunkten entsprechend 64 ms

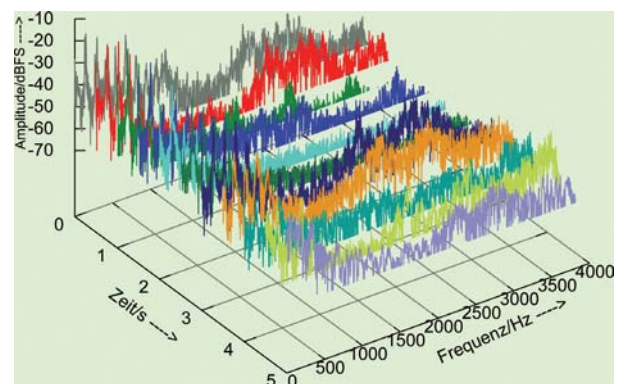


Bild 4: Spektrum im Zeitverlauf einzelner Buchstaben jeweils 500 ms

Klang einer (männlichen) Stimme jedoch erstrangig. Die Betrachtung von S und F zeigt noch etwas Interessantes: Wenn bei einer analogen Sprachmodulation in diesem Frequenzbereich um 1,5 kHz kein gleichmäßiger Amplitudenverlauf von Mikrofon, Mikrofonverstärker oder Sender vorhanden ist, dann beeinträchtigt dies diese Möglichkeit des Hörzentrums, S und F zu unterscheiden. Im Modulationstrimmen erfahrene Operator wissen, dass es auf diesen Bereich besonders ankommt, um eine gute Sprachverständlichkeit zu erzielen. Sie suchen daher nach Mikrofon-Sender-Kombinationen, durch die die Verständlichkeit im wahrsten Sinne hochgezüchtet wird. Erst in zweiter Folge steht der angenehme Klang der ausgesendeten Sprachmodulation. Sie wissen: Nur so kann man im Pile-Up gut durchdringen. Für die Klönrunde haben Sie dann manchmal eine andere Einstellung. Wir wollen die Möglichkeiten des Modulations-Trimmens hier nicht weiter vertiefen.

Spracherzeugung im Sprachtrakt

Das Ergebnis der Spracherzeugung haben wir nun anhand von zwei Beispielen von

Vokalen A und E und den Konsonanten S und F grob analysiert. Um daraus ein Modell zu bilden, lohnt es sich, einen Blick auf die physikalische Spracherzeugung zu tun.

Im **Bild 6** ist der vertikale Schnitt durch einen in den Nacken gestreckten Kopf schematisch dargestellt. Der Schwerpunkt der Darstellung liegt dabei auf dem Stimmapparat. Links ist der Nacken, rechts das Gesicht mit Kinn, Mund- und Nasenraum zu sehen. Der obere Teil des Kopfes ist nicht dargestellt. Zur Orientierung ist der Knochenbau, Wirbelsäule links, Gaumen in der Mitte, sowie die Zunge mit eingezeichnet. Im Einzelnen ist in der Mitte der Mundraum mit Zähnen und Lippen, darüber der Nasenraum und die Nase als Profil zu erkennen. Der Luftraum im Artikulationstrakt ist blau unterlegt dargestellt. Bemerkenswert groß ist der Zungenmuskel – er überdeckt hier die untere Zahnreihe bei den

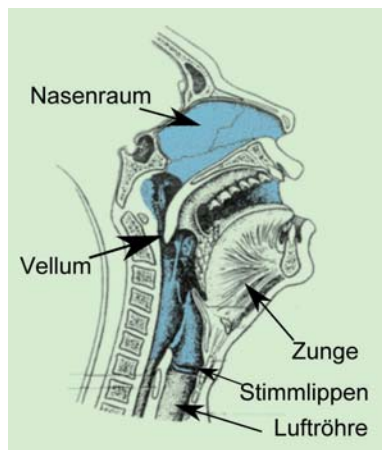
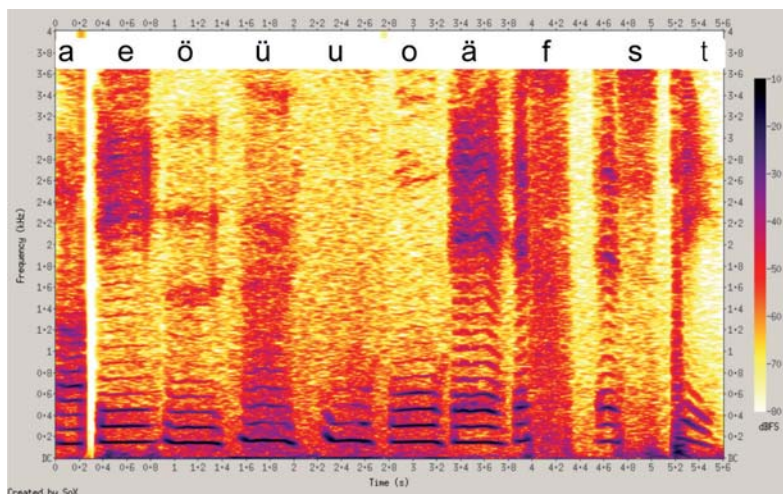


Bild 6: Querschnitt durch den Stimmapparat nach [6]

geöffneten Lippen. Daraus kann ein physikalisches Modell entwickelt werden, welches an [9] angelehnt ist.

Das schematisch dargestellte Modell im **Bild 7** zeigt die Lunge als Speicher für die Luft, der durch variable Muskelkraft einen Luftstrom erzeugen kann. Im rechten Teil des Bildes ist der Artikulationstrakt zu sehen. Er besteht aus Hals, Nasen- und Mundraum. Der Mundraum wird durch Vellum, Zunge, Zähne und Lippen begrenzt, die wiederum durch die Muskelkraft ihre Lage weitestgehend und sehr schnell verändern können. In diesem Modell ist zu erkennen, wie durch Muskelkraft die Lunge über die Luftröhre einen die Intensität der Stimme bestimmenden Luftstrom den Stimmbändern (hier nur die Stimmritze hervorgehoben) zuführt. Diese Stimmbänder sind physikalisch als ein schwingfähiges Feder-Masse-Dämpfer-System modelliert und so schematisch dargestellt. Im Wesentlichen



wird hier durch die unterschiedlichen Muskelspannungen und -kontraktionen impulsförmig die Grundschwingung erzeugt.

Die Zerlegung des Zeitverlaufs nach Fourier weist nach, dass Impulse eine große Anzahl von Oberschwingungen im Frequenzgang enthalten. Durch den daran anschließenden Artikulationstrakt, den mit sehr unterschiedlichen, wiederum durch Muskeln steuerbaren Querschnitten und Längen ausgestattete Vokaltrakt mit einem Teiler für Mundraum und Nasenraum, entsteht physikalisch gesehen ein veränderlicher Resonanzraum, der wie ein Filter wirkt und mit dem die vorhandenen Oberschwingungen aus der impulsförmigen Erzeugung hervorgehoben oder abgedämpft werden können. Bei der Spracherzeugung werden in diesem Vokaltrakt die Weite und Länge und die unterschiedlichen Resonanzräume zeitlich ständig verändert. Damit können wir die einzelnen Phoneme der Sprache erzeugen.

Codierungsverfahren für eine digitale Sprachübertragung

Als nächstes wollen wir der Frage nachgehen, welche minimale Informationsmenge digital zu übertragen ist, um eine Sprachübertragung zu ermöglichen. Würde man alle persönlichen Merkmale der individuellen Stimme des Sprechers herausnehmen, käme man zu einer Textübertragung, deren Informationsmenge bei einigen 10 bis wenige 100 Bit pro Sekunde liegt. Spracherkennungs- und Diktiersysteme leisten dieses inzwischen recht zuverlässig auch bei normaler Sprechgeschwindigkeit. Dann wäre lediglich eine Textübertragung notwendig. Auf der Empfängerseite würde dann aus dem Text mit einem Text-to-Speech (TTS)-System wieder eine Stimme er-

zeugt. Waren es früher Roboterstimmen, so können heute moderne TTS-Systeme einen ermüdungsfrei zu hörenden, angenehmen Klang erzeugen. Diesem Klang fehlen allerdings die persönlichen Merkmale des Sprechers. Philipp Reis, Erfinder des Telefons, hatte bereits um 1860 mit einer 1-Bit-Amplitudenquantisierung Sprache übertragen können, wo zumindest noch der Sinn des gesprochenen erkennbar war, zuhören möchte man da allerdings nicht.

Die modernen TTS-Systeme entwickeln zu dem Text eine eigene Betonung. Diese muss nicht unbedingt mit der des Sprechers übereinstimmen. Immerhin, für den sehr eingeschränkten Vokabelschatz bei einem Contestbetrieb würde es wahrscheinlich ausreichen, da uns die Emotionen des Sprechers in der Regel dabei nicht besonders interessieren. Für eine Klönrunde ist das aber wohl inakzeptabel.

Wenn Stimme und Musik übertragen werden sollen, ergeben sich einige, sehr bekannte andere Möglichkeiten für die digitale Codierung. Dazu gehören besonders MP3, OGG-Vorbis, AAC und die im Mobilfunk verwendeten Codierungsverfahren. Auch diese Verfahren erlauben eine Quellencodierung, die die Informationsmenge deutlich vermindern kann. Als lizenzfreie Codiersoftware für Sprachanwendung, die sowohl niedrige Bitraten, als auch hohe Qualität codieren können, sind SPEEX [15] und OPUS [17] zu nennen. Beide Programme können eine sehr gute Verminderung der zu übertragenden Informationen erreichen. Bitraten hinunter bis zu einigen Kilobit pro Sekunde sind möglich. Den für Kurzweile jedoch erforderlichen sehr niedrigen Datenstrom von etwa 1...2 kBit/s können sie nicht hinreichend gut erzeugen.

Grundsätzlich erscheinen jedoch beide Programme durchaus geeignet, wenn man die Verzögerungszeiten (Latenz) hinnimmt und ein Übertragungskanal zur Verfügung steht, der mehr als 4 kBit/s zuzüglich der Fehlerkorrekturdaten erlaubt.

In **Bild 8** sind Bitraten und Verzögerungszeiten verschiedener Codecs dargestellt. Auch der G.729, ein Sprach-Coder für die IP-Telefonie, ist eingeordnet. Deutlich sichtbar in der linken, oberen Ecke des Bildes ist, dass sowohl der G.729- als auch der Opus- und der Speex-Codec sowohl relativ niedrige Bitraten mit einigen Kilobit pro Sekunde, als auch relativ kleine Verzögerungszeiten haben, wie sie für das Wechselsprechen notwendig sind. Sie werden daher mit einer Skala an der rechten Seite der Grafik als Realtime-fähig (live) eingestuft.

Die gewünschte Quellencodierung und Sprachsignalverarbeitung mit sehr kleinen Bitraten für den digitalen Kurzwellenfunk lässt sich damit nicht erreichen. Auch der AMBE-Codec, eingesetzt bei D-Star, scheidet wegen seiner relativ hohen Datenrate mit 2,4 kBit/s deswegen aus meiner Sicht aus. Die im zweiten Artikelteil (CQ DL 2/14) ermittelte praktisch erreichbare Kanalkapazität (6) hat bereits deutlich gemacht, dass wir bei Verwendung der hier dargestellten Codecs nur dann eine digitale Sprachübertragung realisieren können, wenn dafür eine größere Bandbreite als die hier angestrebten 1200 Hz eingesetzt würde.

Prinzip des Vocoders

Vocoder, Quellencodierer für Sprache, (Voicecodec) sind eine recht alte Erfindung [12]. Die modernen Verfahren arbeiten auf der Sendeseite mit einer Rückkopplung, um die Parameter in einem autoregressiven Modell (AR-Modell) des Modells schätzen zu können. Die Schätzung wird als Lineare Prädiktion bezeichnet.

Auch die im Mobiltelefon verwendeten Sprachcoder nach dem GSM-Standard arbeiten nach einem als Linear-Prädiktion-Coder (LPC) bekannten Prinzip, der das Modell der Spracherzeugung aus **Bild 7** übernimmt. Damit entsteht das **Bild 9**. Aus der Grundfrequenzanforderung werden im stimmhaften Zweig-Impulse erzeugt. Die Energie wird mit einem Verstärker/Multiplizierer eingestellt und in der Sprachtaktmodell (das Filter) aus dem Mund gesprochen. Im stimmlosen Zweig wird ein Rauschgenerator verwendet (ausströmende Luft).

Die wesentlichen Parameter des Modells sind

- stimmhaft oder stimmlos,
- Grundfrequenz des Impulsgenerators,
- Parameter des Filters,
- Energie.

Die Parameter dieses Modells werden in der Regel nach einigen Millisekunden – typisch 20 ms – erneut geschätzt und auf der Sende- und Empfangsseite neu eingestellt. Aus der Beobachtung kann man ableiten, dass für diesen Zeitraum von einigen ms diese Parameter hinreichend stabil sind. Zumindest bei schwieriger Verständigung wird der Sprecher immer zu langsamer Sprache neigen. Dafür gilt die gerade geforderte Stabilität der Parameter in jedem Falle. Es gibt verschiedene auf LPC beruhende Klassen von Sprachcodern [2]:

Die **Waveform-Codierung** überträgt nicht nur die Modellparameter, sondern zusätzlich die Differenz zwischen dem Modell und dem ursprünglichen Signal. Das Residual-Werteübertragungsverfahren wird daher RELP abgekürzt. Auch der GSM-Standard setzt darauf auf. Hiermit lässt sich relativ hohe Qualität erzeugen, da das Anregungssignal sich auf der Empfängerseite wieder gut rekonstruieren lässt.

Die **Codebook-Anregung** setzt dem Prädiktionsfehler mit einem hoch effizienten Vektorquantisierer um. Dabei werden bestimmte Abschnitte in einem Codebuch stellvertretend für das Muster abgelegt. Dieser Typ ist als CELP-Verfahren (Code-Excited LPC) bekannt. Die Vektorquantisierung führt zu Abweichungen zwischen dem Original und dem Übertragenen. Es ist notwendig, dieses Codebuch auf der Sende- und auf der Empfängerseite vorzuhalten.

Die **Pitch-Anregung** ist besonders effizient, weil das Anregungssignal bei dieser Klasse nicht mehr übertragen wird. Es ist das am weitesten gehende Modellverfahren. Bei diesem extrahiertem Verfahren werden nur noch die Parameter übertragen. Für eine möglichst gute Wiedergabe ist daher die Schätzung der Parameter von großer Bedeutung. Mit dieser Klasse können jedoch die geringsten

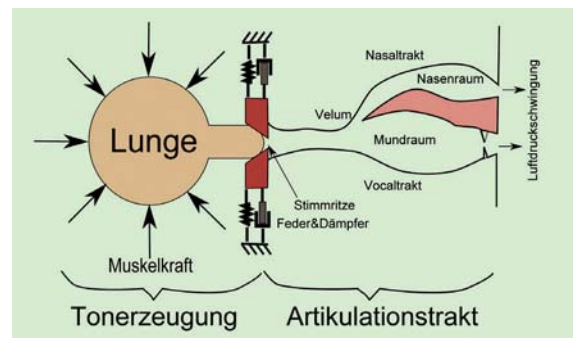


Bild 7: Modell des Sprachtraktes nach [9]

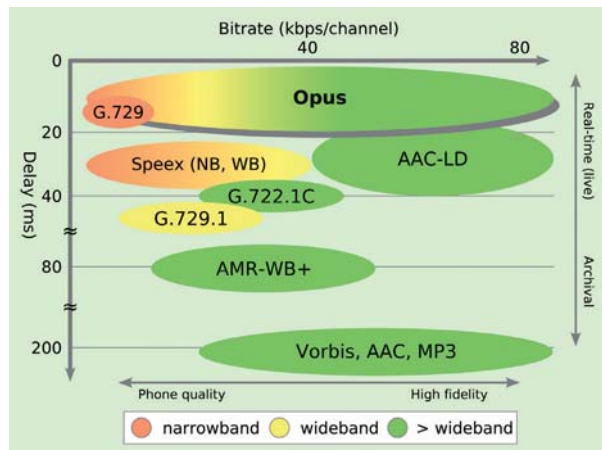


Bild 8: Latenzen und Übertragungsraten für verschiedene Sprach- und Musik-Codecs nach [17] bei verschiedenen Nutzbandbreiten

ten Bitraten erreicht werden. Zumindest der Codec2 [10] gehört zu dieser Klasse. Er verwendet das im Bild 9 gezeigte Modell weitestgehend.

Einführung digitaler Sprachübertragung im Afu

Wenn man im Amateurfunk ein neues Verfahren etablieren möchte, so lohnt es sich danach zu schauen, dass die Schwelle für die Anwendung möglichst niedrig ist. Das gilt sowohl für die finanzielle als auch die apparative Seite. Es muss ferner einfach und preiswert sein. Daher haben sich bei Experimenten in der Vergangenheit verschiedene Verfahren besonders einfach dadurch einführen lassen, dass diese als Programme für den Stationsrechner zur Verfügung standen. Wurden in den früheren Jahren noch Modems gelötet, so werden diese heute durch den Stationsrechner oder einen kleinen Mikrocontroller mit einem Signalverarbeitungsprogramm und einer Soundkarte realisiert. So könnte sich auch die digitale Sprachübertragung unter Funkamateuren verbreiten – einfach soll es sein, kosten soll es auch nichts. Im Folgenden wollen wir uns die unterschiedlichen, bereits im Amateurfunk bekannten Verfahren noch einmal näher anschauen und deren mögliche Hinderungsgründe für die große Verbreitung herausfiltern.

Das für den Kurzwellenbetrieb schon seit längerem bekannte Programm WinDV nutzt einen patentierten Vocoder aus der Klasse CELP. Das Verfahren lässt sich damit nicht aus diesem Programm herauslösen, geschweige denn in einer anderen Hardware realisieren, ohne die Patentrechte zu berühren.

Auch andere Vocoder sind meist patentiert oder beruhen auf patentierten Verfahren, sodass eine Verwendung außerhalb des vorgesehenen Einsatzes schwierig, oder rechtlich unmöglich wird. Ein gutes Beispiel dafür ist der AMBE-Vocoder. Er erzeugt typisch 2400 Bit/s in einem Hardwarechip. Laut Datenblatt sind zwischen 2000 und 9600 Bit/s einstellbar. Die Algorithmen zur Schätzung der Parameter sind Gegenstand des Patents für diesen Chip und auch nicht genauer bekannt. Mehr noch: Sie unterliegen allein dem Hersteller, dieser kann sie jederzeit ändern. Selbst die genaue Bedeutung der Bits im Sprachdatenframe ist nicht offen zugänglich.

Gegenüber kommerziellen Funkanwendungen haben wir im Amateurfunk andere Anforderungen. Diese laufen

steigen. Das gilt auch für die speziellen Zusatzgeräte für Kurzwellenanwendung (AOR/ARD9800).

Eine einfache Anpassung der Hardware an neue Algorithmen für eine bessere Sprachcodierung ist meist nicht möglich. Genau damit sehen die frühen Käufer (early adopters), dass sich ihre Investitionen nicht lohnen – sie haben dann immer alte und teurere Geräte. Daher sollten Verfahren für den Amateurfunk ein kostenloses Update auch nach langer Zeit noch ermöglichen, um an der aktuellen Entwicklung teilzunehmen – der Amateurfunk ist immer experimentell. Komplexe Algorithmen, in Hardware realisiert, erfordern auch mit heutiger Technik bei geringen Stückzahlen immer wesentlich höhere Anfangsinvestitionen und das ist dann immer eine hohe Schwelle. Somit einfach eine hohe Schwelle für die experimentelle Weiterentwicklung.

Bekanntes Verfahren der digitalen Sprachübertragung

Leider lassen auch die Lösungen mit patentierten kommerziellen, in Software ablaufenden Codecs, wie dem CELP

Bit/s FEC-Daten mit 36 Trägern. Bei der Spezifikation von G4GUO ist eine Synchronisierzeit von etwa 1 s vorgesehen. Gelingt die Synchronisation nicht, oder kommt jemand erst später auf die Frequenz, wie beim KW-Betrieb durchaus nicht ungewöhnlich, so sind die Sprachdaten nicht zu decodieren. Schnelles Wechselsprechen wird dadurch wesentlich beeinträchtigt und der fehlende genau festgelegte Kanalfrequenzplan ist ein erhebliches Hindernis – das wird wohl auf KW auch nicht kommen.

HamDRM/WinDRM [8] eignet sich wohl grundsätzlich, scheint aber doch sehr eingeschränkt einsetzbar und nur ablauffähig auf Stationsrechnern, die mit Windows betrieben werden können. Die Funkübertragung mit WinDRM erfordert einen besseren Signal-/Störabstand und mit 2,3 oder 2,5 kHz Bandbreite fast eine volle SSB-Kanalbreite. Das Programm nutzt den Speex-Codec mit im Mittel 2400 Bit/s. Leider erfordert es auch einige Sekunden Synchronisierzeit, sodass ein im Amateurfunk typisches, schnelles Wechselsprechen sich so nicht erreichen lässt.

Es zeigt sich, dass einige Verfahren teilweise schon mehr als zehn Jahre existieren. Die genannten Verfahren haben überwiegend mit 2400 Bit/s Sprachdaten und wollen die Bandbreite des SSB-Kanals praktisch vollständig ausnutzen. Aus dem praktischen Kurzwellenbetrieb weiß der erfahrene KW-Funkamateur, wie selten er die volle Bandbreite eines SSB-Kanals finden und behaupten kann. Die Verfahren weisen daher im praktischen Kurzwellenbetrieb erhebliche Einschränkungen durch die unvermeidlichen Störungen auf. Erste Ansätze zeigen nun, dass es eventuell gelingen könnte, mit einer geringeren Übertragungsrate und damit auch mit einer deutlich geringeren Bandbreite als dem vollen SSB-Kanalauskommen. Einzelne hier genannten Verfahren basieren auf teilweise langen Synchronisationszeiten für die digitale Sprachübertragung und der Ableitung von anderen Verfahren, die nicht für schnelles Wechselsprechen entwickelt wurden, z.B. Rundfunk. Das macht deren Einsatz im Kurzwellenamateurfunk allerdings sehr aufwändig und schwierig.

Im weiteren wollen wir also schauen, welche guten Ansätze für ein schnelles Wechselsprechen heute schon möglich sind.

(wird fortgesetzt)

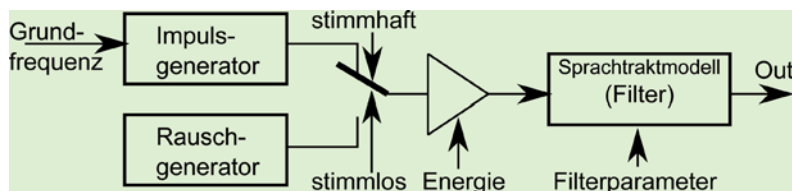


Bild 9:
Quelle-Filtermodell
der Sprach-
erzeugung

eher auf eine möglichst langlebige und auf eine möglichst kompatible Verwendung von Geräten hinaus. Wegen des Hobbycharakters spielt häufig auch der Preis eine wichtige Rolle. Die Implementierung in Hardware (AMBE-Chip) hat zudem aus Sicht des Hobbyisten den Vor- oder Nachteil, dass hier bei der Weiterentwicklung ein neues Gerät zu erwerben ist. Das Beispiel D-Star zeigt, dass nicht nur in einem Funkgerät integrierte Lösungen, sondern auch einfache Zusatzgeräte, oder sogar Selbstbau mit dem Hardware-Chip denkbar sind. Zusatzgeräte für die auf VHF und UHF verwendeten Handfunkgeräte sind wegen der Größe unpraktisch und daher nicht verbreitet.

Im Stationsbereich, und besonders bei den zumeist stationären Kurzwellenanlagen, ist die Verwendung von Zusatzgeräten leichter möglich und üblich. Diese Zusatzgeräte sind in der Regel relativ teuer und können auch schon mal den Preis eines einfachen Funkgerätes über-

oder MELP, kaum eine Aktualisierung zu. Zusätzlich sind diese weder frei verwendbar noch haben sie besonders hervorstechende Eigenschaften für die Anwendung bei der Kurzwellenübertragung. Der MELP-Vocoder lässt sich allerdings bis hinunter zu 600 Bit/s betreiben. Der TWELP-Vocoder-Algorithmus [7] ist ein neuerer Ansatz der ebenfalls verspricht, bis hinunter zu 600 Bit/s zu erreichen. Es ist wiederum ein geschlossenes Verfahren. Das AOR-Verfahren ARD9xxx hat zwar die prinzipielle Eignung für Kurzwelle, ist aber nicht nur relativ teuer, sondern auch noch sehr wenig verbreitet. Das Protokoll und die zur Grundlage des Gerätes verwendeten Prinzipien sind von G4GUO entwickelt worden und in [1] erläutert. Das in den Geräten verwendete Protokoll scheint nicht hinreichend offen gelegt zu sein. Das System verwendet praktisch die volle Bandbreite des SSB-Kanals von 0,3...2,4 kHz und sendet 2400 Bit/s Sprachdaten sowie 1200

Das im Weiteren einzeln beschriebene und daraus zusammengesetzte Verfahren kann viele dieser offenen Fragen lösen. Kern dieses hier beschriebenen Verfahrens ist der weiterentwickelte und offene, sowie patentfreie Vocoder.

Grundlage einer neuen Sprachübertragung über KW

Prof. Dr. Michael Hartje, DK5HH

Aufbau und Funktion von Codec2 ist Gegenstand des vierten Teils der Beitragsserie. Von David Rowe, VK5DGR, entwickelt, eignet er sich in besonderem Maße für den Einsatz auf Kurzwelle.

Der von David Rowe, VK5DGR, unter Mitwirkung weiterer Funkamateure entwickelte Algorithmus des Codec2-Vocoders gehört in die Klasse der Pitch-Anregung nutzenden hoch effizienten Vocoder. Das Besondere an Codec2 ist nun, dass die Parameterschätzung und deren Vektorquantisierung optimiert sind für eine leider vorkommende, fehlerhafte Übertragung, wie sie zum Beispiel bei Kurzwelle entsteht. Dabei wird das Ziel erreicht mit Codec2, bei der Rekonstruktion des gesendeten Signals aus dem fehlerhaft übertragenen binären Signal möglichst geringe Abweichungen und vor allen Dingen möglichst wenig Artefakte zu erzeugen.

Vocoder werden üblicherweise auf eine möglichst gute Rekonstruktion des Sprachsignals ausgelegt. Eine gleichzeitige Berücksichtigung der Eigenschaften des Übertragungsweges ist hier eher unüblich. Die vorausschauende konstruktive Berücksichtigung der Übertragungsfehler einer typischen Kurzwellenübertragung zeichnet den Codec2 aus. Das entstandene Modell dieses Vocoders ist im **Bild 10** dargestellt: Die abgetastete Sprache wird in einer Pitch-Schätzung (Grundfrequenzschätzung), einer Fouriertransformation und in einer Parameterschätzung des Filters (LPC) bearbeitet. Die geschätzten Parameter werden anschließend als diskrete Zahlen quantisiert. Eine Schätzung gilt für einen Zeitraum von 20 ms. Die entstandene Informationsmenge ist also später binnen 20 ms in einem Rahmen (Frame) zu übertragen.

Geringe Bitraten

Mit den aktuellen Versionen des Codec2 kann die Verdichtung auf 1200...1400 Bit/s vorgenommen werden. Höhere

Bitraten mit 2400 und 3200 Bit/s sind möglich und erlauben eine bessere Sprachqualität. Der Vocoder soll lediglich Sprache übertragen. Das bringt den Vorteil mit, das Hintergrundgeräusche nicht codiert und nicht übertragen werden, weil sie nicht in das Muster/das Modell der Spracherzeugung passen. Der Vocoder mit diesem Modell erzeugt je nach der gewählten Bitrate sehr unterschiedliche Frames. Das folgende Datenframe in **Tabelle 1** ist für den 1400-Bit/s-Modus. Dieses Datenframe wird sich mit der Weiterentwicklung des Codec2 möglicherweise noch wieder ändern. Das gesamte Frame mit den vier Teilframes wird bei der Version für 1400 Bit/s für 40 ms bestimmt; es werden also 25 Frames f mit $b = 56$ Bit pro Sekunde übertragen.

$$D = f \cdot b = 25 \text{ Frame/s} \cdot 56 \text{ Bit/Frame} = 1400 \text{ Bit/s} \quad (10)$$

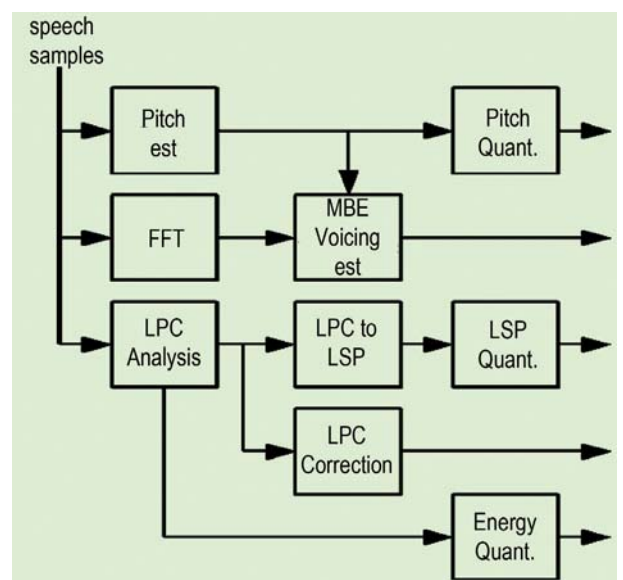
Die jeweils abbildenden Teilframes F1 und F3 enthalten nur jeweils 1 Bit für den Parameter stimmlos/-haft. Diese Bits werden im Teilframe F2 und F4 mit untergebracht – die Informationsmenge wird in F2 für F1 und F2 gemeinsam in 2 Bit untergebracht. Die spektralen Amplituden werden nur einfach je 40 ms in F4 übermittelt, dann allerdings mit hoher Auflösung. Da diese Parameter

innerhalb 40 ms sich nicht grundsätzlich ändern können, reicht diese Einstellung aus.

Die spektralen Amplitudenparameter werden durch so genannte spektrale Linienpaare (LSP) abgebildet, die gegenüber den originalen Filterkoeffizienten (LPCs) bei der Quantisierung etwas unanfälliger gegenüber Quantisierungsfehlern sind – eine übliche Anwendung bei Sprachcodern, z.B. bei Speex. Sowohl die spektralen Amplitudenparameter (LSPs), als auch die Energie und die Grundfrequenz (Pitch) werden durch eine verlustbehaftete nichtlineare Quantisierung (Vektorquantisierung und Codebuch) auf sehr wenige Bits abgebildet.

Entwicklung geht weiter

Es gibt weiterhin Überlegungen, das Verfahren noch dahin zu entwickeln, noch weniger Parameter zu aktualisieren oder deren Gültigkeit für das Modell zu strecken bis auf 80 ms. Dies hätte zur Folge, dass noch niedrigere Bitraten erreicht werden können. An dieser Stelle ist die Entwicklung noch nicht abgeschlossen. Das ist allerdings sehr zeitaufwändig und mit vielen subjektiven Tests verbunden, um die Qualität der Sprachübertragung bei diesen neuen Formaten



Den Autor erreichen Sie unter Prof. Dr. Michael Hartje, DK5HH Nettelbeckstr. 28 28201 Bremen dk5hh@dar.c.de

Die Literaturangaben finden Sie in Ausgabe 2/14, S. 10

Bild 10: Modell des Codec2-Vocoders

auf hohem Niveau zu halten. So ist die Entwicklung auf Beteiligung aus dem Amateurfunk angewiesen. Die Lösung in Software erlaubt allerdings später ein einfaches und kostenloses Upgrade. Insgesamt ist bis heute ein äußerst flexibles System entstanden, bei dem Sprache durch geschätzte Modellparameter auf Bitraten zwischen 1200...3200 Bit/s umgesetzt werden kann. Der Ablauf dazu soll weiter unten gezeigt werden.

FEC – vorausschauende Fehlerkorrektur

Anders als bei den im Amateurfunk gut bekannten Datenübertragungen (z.B. Packet Radio) ist es bei einer Sprachübertragung nicht möglich, eine fehlerhafte Datenübertragung nach Fehlererkennung zu wiederholen.

Sprachübertragung ist eine Echtzeitanwendung; es verbleibt keine Zeit für die Wiederholung der Informationen. Damit fällt das einfache Verfahren mit der Fehlererkennung durch Bilden einer Prüfsumme und dem Vergleich mit einer übermittelten Prüfsumme aus. Die bei Ungleichheit entdeckte fehlerhafte Übertragung kann aus Zeitgründen nicht korrigiert werden.

Um aber die auftretenden Übertragungsfehler abzusichern, wird ein anderes Verfahren angestrebt. Eine sichere Übertragung der mit der Quellencodierung gewonnenen geringen Informationsmenge des gesamten Sprachinhaltes wird meist mit Blick auf die bei der Übertragung entstehenden Fehler um eine vorausschauende Fehlerkorrekturmöglichkeit erweitert. Dies erfordert zunächst die möglichst genaue Kenntnis über die entstehenden Übertragungsfehler.

Dann werden dafür zusätzliche Informationen mit einem speziellen Rechenverfahren zu dem Paket hinzugefügt, sodass dann später mit der Kenntnis dieser Rechenvorschrift fehlerhaft übertragene Informationen vollständig korrigiert werden können.

In der Praxis muss vorher geklärt werden, wie gut die zu übertragenden Informationen gegen deren Verfälschung bei der Übertragung abgesichert werden müssen. Je besser die Absicherung, desto umfangreicher werden die hinzuzufügenden Informationen und die Rechenvorschrift.

Das **Bild 11** zeigt das gesamte System im Überblick. Das Sprachsignal (links oben) mit einem dicken Datenstrom wird in der Quellencodierung, z.B.

Codec2, von Redundanz und Irrelevanz befreit und in einen dünnen Datenstrom codiert. Im nächsten Schritt erfolgt eine Kanalcodierung, die die erwarteten Übertragungsfehler durch geschickte Rechenverfahren um zusätzliche Informationen zur Fehlerkorrektur erweitert – ein dickerer Datenstrom entsteht. Im Weiteren wird mit einer Leitungscodierung, das ist typischerweise ein Modem, eine Umsetzung der Daten in übertragbare Signale vorgenommen.

Bei der Übertragung werden Störungen additiv und multiplikativ hinzugefügt. Additive Störungen sind zum Beispiel Splatter vom Nachbarkanal oder einfach nur Rauschen. Multiplikative Störungen sind zum Beispiel Fading. Sie verändern die Amplitude des Signals und besonders auf Kurzwelle kann das sogar bei sehr kleinen Bandbreiten sehr stark frequenzselektiv wirken.

Auf der Empfängerseite wird dann alles rückwärts bearbeitet. Mit der Leitungscodierung entstehen im Modem aus den Tönen binäre Signale. In der Kanaldecodierung werden nun die entstandenen Übertragungsfehler in den Datenpaketen bestimmt und es wird versucht, diese Fehler zu korrigieren. Daran schließt sich die Quellendecodierung an, bei den aus dem inzwischen sehr, sehr dünnen Datenstrom wieder die beim Sender entnommene Redundanz hinzugefügt wird, sodass daraus eine ausreichende Menge von Abtastwerten berechnet werden kann, die bei einer Tonausgabe auf dem Lautsprecher schließlich die Sprache des Sprechers erklingen lässt.

Trx als gewöhnliches Gerät

Das **Bild 12** überträgt das grundsätzliche Konzept aus Bild 11, welches für jede digitale Sprachübertragung auch auf Leitungen Anwendung findet, auf eine Funkübertragung. In dem Bild 12 ist der Weg der Sprache vom Mikrofon des Senders zum Lautsprecher des Empfängers mit den Blöcken A/D-Umsetzung, Quellencoder (Codec encoder), Leitungscoder (FEC encoder) und Modulator im Sender dargestellt. Wir erinnern uns: Der Sender ist ein vorhandener SSB-KW-Trx. Das Konzept lässt sich natürlich auch auf VHF übertragen. Allerdings treten dort die Störungen bei der HF-Übertragung häufig mit anderen Eigenschaften auf.

Die vorausschauende Fehlerkorrektur (FEC encoder) muss die Eigenschaften der HF-Übertragung und die dabei

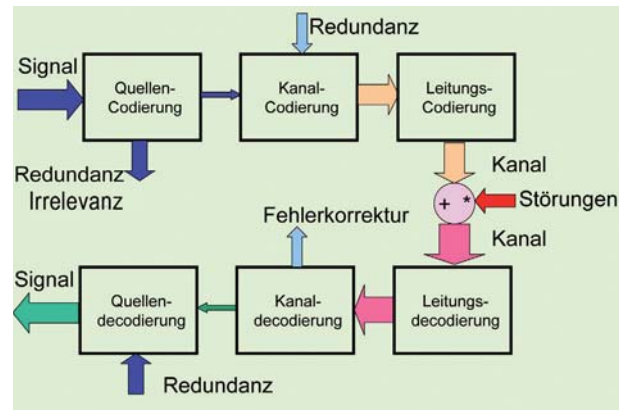


Bild 11: Konzept digitale Verarbeitung bei digitaler Übertragung

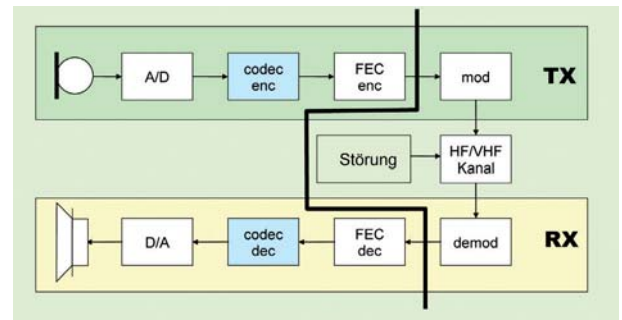


Bild 12: Digitale Verarbeitung bei der Funkübertragung

möglicherweise entstehenden Fehler mit ihren Eintretenswahrscheinlichkeiten für das übertragene Datenpaket kennen. Nur dann ist es möglich, entstandene Übertragungsfehler fehlerlos zu korrigieren. Aus diesem Umstand heraus muss nicht nur das Modem, sondern auch die Forward-Error-Correction für die Kurzwellenübertragung andere Eigenschaften aufweisen, als für VHF/UHF.

Auch aus diesem Grunde erscheint eine einfache Übernahme von D-Star für Kurzwelle nicht besonders sinnvoll. Da die Forward-Error-Correction jedoch Bestandteil des AMBE-Chip ist, dürfte es nicht einfach sein, diesen mit den entsprechenden Eigenschaften der Kurzwellenübertragung zu parametrieren – wahrscheinlich ist dies lediglich dem Hersteller des AMBE-Chips möglich.

Auf der Empfängerseite wird das gesamte Verfahren wieder rückwärts in den gleichen Schritten durchgearbeitet. Um die zusätzlich bei der Kanalcodierung erforderlichen Informationen möglichst gering zu halten, nehmen die im FreeDV und bei Codec2 als Einzelprogramme verfügbaren Routinen für die Forward-Error-Correction nur einen Ausschnitt aus dem Frame. Nur einige, ausgewählte, wichtige Bits werden auf diese Weise geschützt. Dazu gehören die beiden stimmlos/-haft-Bits jedes Halbframes sowie die Codierung der Grundfrequenz. Die Forward-Error-

Correction-Rechenvorschrift muss also sowohl Kenntnisse davon haben, welche Bits innerhalb eines Frames besonders wichtig für eine fehlerfreie oder möglichst störungsarme Sprachübertragung sind, als auch die genauen Kenntnisse über die bei der Übertragung entstehenden Fehler dieser geschützten Bits haben.

Rechen-Vorschrift zur Decodierung

Diese Rechenvorschrift kann bei der Decodierung auch die Anzahl der Fehler bestimmen. Wenn zu viele Fehler auftreten, würde die spätere Sprachdecodierung zu Fehlern oder Artefakten führen. Daraus lässt sich eine weitere Möglichkeit ableiten, mit der bei zu vielen erkannten Fehlern, z.B. das letzte fehlerfreie Frame wiederholt wird, oder der Ausgang stumm geschaltet werden kann. Beides ist heutzutage bei digitalen Übertragungssystemen für Sprache und Bild eingesetzt (bei DVB-T entstehen daraus die Bilder mit den großen Farbpixeln – das Klötzchenbild).

Bei dem Programm FreeDV [4] gibt es zwei, in der grafischen Bedienoberfläche wählbare Betriebsmodi:

- 1400 Bit/s Sprachdaten-Codec ohne zusätzliche Forward-Error-Correction. Dies führt zu einer ungeschützten Übertragung der Sprachdaten. Dieser Modus 1 war in der ersten veröffentlichten Version und wird heute eher aus historischen – Dezember 2012! – Erwägungen weiterhin zur Verfügung gestellt.
- 1300 Bit/s Sprachdaten-Codec mit zusätzlicher Forward-Error-Correction von 300 Bit/s. Daraus entstehen 1600 Bit/s zu übermittelnde Daten. Der Sprachcodec codiert mit 1300 Bit/s geringfügig anders, als der Sprachdaten-Codec für 1400 Bit.

Beide Modi lassen sich sehr einfach vor Ort durch Aufrufe der Programme „c2enc“ und „c2dec“ mit dem Parameter „1300“ oder „1400“ in einem Simulationsmodus prüfen. Der Unterschied ist praktisch nicht hörbar.

In der aktuellen Version des Programms FreeDV ist daher der zweite Betriebsmodus mit 1600 Bit/s voreingestellt und wird in der Praxis heute von den meisten KW-Funk-Anwendern dem Modus 1 mit 1400 Bit Sprachdaten ohne Fehlersicherung vorgezogen.

Die **Tabelle 2** zeigt, welche Sprachpaketbits beim 1300 Bit/s-Modus entstehen und welche davon mittels einer Fehlersicherung (FEC) gesichert werden. Die Tabelle entstammt dem Quellcode des FEC-Programmes. Das Verfahren der Fehlersicherung ist nach dem Golay-Algorithmus (23,12) ausgewählt [16]. Dieser ist in der Lage, für die gesicherten zwölf Informationsbits mit elf zusätzlichen, also 23 Bits insgesamt, drei Bitfehler korrigieren zu können (D-Star verwendet ein ähnliches Verfahren). Die ursprünglichen zwölf gesicherten Bits werden mit ebenso vielen Bits aus den 23 Fehlersicherungsbits ausgetauscht und die restlichen 11 Bits des 23 Bit umfassenden Golay-Paketes werden zusätzlich als ein 12-Bit-Paket zu den 52-Bit-Sprachdaten hinzugefügt. Es entstehen damit 64 Bit für jeweils 40 ms für die Übertragung. Genau genommen ist ein Bit in dem zusätzlichen 12er-Paket nicht verwendet.

FDMDV-Modems im Detail

Mit Blick auf das Bild 11 und 12 verbleibt noch, die Leitungscodierung, das Modem, näher zu betrachten. Zuvor wurde schon erwähnt, dass es einen

Unterschied zwischen den Eigenschaften der HF-Übertragung (HF-Kanal) auf Kurzwelle und den Kanaleigenschaften auf VHF und UHF gibt. Geht man davon aus, dass frequenzselektives Fading eine typische Erscheinung für die Kurzwellenübertragung ist, so drängt sich förmlich auf, die Informationsmenge eines Sprachdaten-Frames, zuzüglich zur Fehlersicherung in der gewünschten Bandbreite auf mehrere parallele Träger mit sehr langsamen, einzelnen Symbolzeiten parallel zu verteilen.

Bei sehr langsamen Symbolzeiten können schnelle Veränderungen durchlaufen ohne dass der Symbolinhalt auf der Empfangsseite falsch interpretiert wird. Schnelle Veränderungen mit weniger als 20 ms können z.B. ein Knacken und Splatter aus dem Nachbarkanal sein.

Diese Mehrträger-Modulationsverfahren werden bei vielen modernen digitalen Übertragungsaufgaben eingesetzt. Beispiele gibt es dazu auch im Amateurfunk: MT63, Olivia, aber auch der digitale Hörfunk mittels DRM und viele weitere Anwendungen. Im Amateurfunk gibt es zudem das WinDV für die digitale Sprachübertragung auf Kurzwelle, welches ebenfalls dieses Mehrträgerverfahren verwendet.

WinDV dient als Grundlage

Das Modem von WinDV ist Grundlage für das Modem bei FreeDV. Es wurde vollständig neu programmiert und ist nun frei als Quelle FDMDV verfügbar. Im Programm FreeDV wird es zur Übermittlung der Sprachdaten von Codec2 verwendet. Die Bezeichnung FDMDV leitet sich aus der englischen Bezeichnung für „Frequency Division Multiplex for Digital Voice“ ab.

Im Softwarepaket Codec2 kann das enthaltene, nachprogrammierte Modem, als „fdmdv_enc“ und „fdmdv_dec“ mit Parametern aufgerufen werden. Beim Programm FreeDV sind diese Modemprogramme über deren Bibliotheksfunktionen in das Programm FreeDV einbezogen worden. Das Konzept des Modem-Sendeteils ist im **Bild 13** dargestellt und soll kurz beschrieben werden. Im Beispiel sind hier 1600 Bit/s als Eingangsdatenstrom angenommen. Die 1600 Bit kommen mit den logisch zusammengehörenden Bits in Paketen von 2×32 Bit an. Jeweils 2 Bit werden im Mapper zusammengefasst und auf 16 Träger parallel mit DQPSK aufmoduliert. Alle bilden zusammengenommen ein 32-bittiges-Symbol. Es werden 50 Symbole/s gesendet.

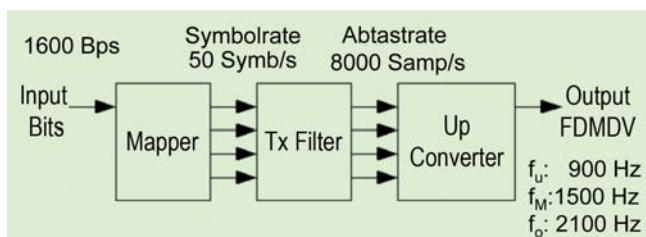


Bild 13: Konzept Modem FDMDV

Tabelle 1

Eigenschaft	Bit in F2	Bit in F4	Bit/frame
Spektrale Amplituden (LSPs)	0	36	36
Energie und Grundfrequenz (Pitch)	8	8	16
Stimmhaft/-los für jeweils 10 ms	2	2	4
Summe Datenbits pro Frame	36	17	56

Tabelle 1: Datenframe des Codec2 in der Version für 1400 Bit/s mit vier Frames zu je 10 ms Länge F1–F4

Tabelle 2

Bedeutung im Paket	Sprach-Bits in F4	gesicherte FEC-Bits
Stimmhaft/-los für jeweils 10 ms	4	4
Grundfrequenz (Pitch)	7	4(MSB)
Energie	5	4(MSB)
Spektrale Amplituden (LSPs)	36	17
Summe im Frame	52	12

Tabelle 2: 40 ms-Frame bei 1300 Bit/s mit 300 Bit/s Fehlerkorrektur. Ein 40-ms-Frame umfasst 64 Bit/s

Um die bei DQPSK-Modulation entstehenden Seitenbänder so zu verringern, dass der Nachbarträger nicht beeinflusst wird, wird je Träger ein digitales Filter (Roll-Out-Filter) zur Impulsformung auf die Abtastwerte aufgesetzt; 16 Filter parallel. Schließlich werden die 16 einzelnen Datenströme mit einer Abtast-rate von 8000 Samples/s mit einem Modulator auf 16 verteilte Frequenzen im Bereich von 900...2100 Hz mit einem Abstand des Modems von 75 Hz im Normalmodus umgesetzt.

Der Abstand ist damit etwas größer, als er bei OFDM [11, 2, 3] minimal (50 Hz) sein müsste. Mit Aufrufparametern kann das Modemprogramm bis zu 20 Träger sowie einen größeren Frequenzabstand erzeugen. Im FreeDV ist ein „wide“-Modus implementiert, bei dem der Abstand der 16 Träger auf 125 Hz angehoben wird. Schließlich wird in der Mitte ein Pilotton erzeugt. Für die Bestimmung der genutzten Bandbreite ist nicht nur der Frequenzabstand und die Zahl der Träger, sondern auch die Seitenbänder der beiden äußeren Träger mit der einfachen Symbolrate s anzusetzen. Dann wird eine Bandbreite b_{wide}

$$b_{wide} = (n + 1) \cdot \Delta f + s = 17 \cdot 125 \text{ Hz} + 50 \text{ Hz} = 2175 \text{ Hz} \quad (11)$$

von 413...2587 Hz. Dieser Frequenzbereich wird bei den meisten SSB-Sendern hinreichend linear übertragen. Das ist natürlich in der Praxis zu prüfen, ob der vorgesehene SSB-Sender dieses auch erlaubt. Mit den 16 Trägern und im Normalmodus mit $f = \Delta 75 \text{ Hz}$ entsteht eine Bandbreite b_{normal}

$$b_{normal} = (n + 1) \cdot \Delta f = 17 \cdot 75 \text{ Hz} + 50 \text{ Hz} = 1325 \text{ Hz} \quad (12)$$

Mit dem alten 1400 Bit-Modus werden nur 14 Träger verwendet, sodass die Bandbreite sich noch weiter vermindert auf

$$b_{1400} = n \cdot \Delta f + s = 15 \cdot 75 \text{ Hz} + 50 \text{ Hz} = 1175 \text{ Hz} \quad (13)$$

Im Empfangsfall kann man anhand der Bandbreite des Signals den entsprechenden Modus unterscheiden und derzeit manuell auswählen.

32 Bit in 20 ms übertragen

Ein Symbol mit seinen 32 Bit wird nun innerhalb von 20 ms übertragen. Es ent-

hält die Informationsmenge für ein halbes Sprachdaten-Frame inklusive der Fehlersicherungs-Informationen.

Zwei halbe Sprachdaten-Frames zusammen werden binnen 40 ms übertragen.

Erst mit einem neueren Ansatz werden die 64 Bit Sprach- und Fehlersicherungsbits vor der Über-

gabe an das Modem verwürfelt (Interleave). Das Interleave ist im Bild nicht explizit dargestellt – es ist derzeit bei FreeDV nur für den 1600-Bit-Modus vorgesehen. Andere Modi lassen sich allerdings leicht implementieren, fehlen derzeit aber noch. Das demodulierende empfangsseitige (De-)Modem ist im **Bild 14** dargestellt.

In einem Downconverter werden die DQPSK modulierten Träger einzeln heruntergemischt. Sie durchlaufen jeweils einen Empfängerfilter und werden decodiert. Dazu ist eine Taktregenerierung erforderlich.

Diese ermöglicht, das einschwingende Symbol eines Trägers mit dem die Information repräsentierenden Zuständen 0, 1, 2 oder 3 zum richtigen Zeitpunkt, etwa in der zeitlichen Mitte des Symbols, abzutasten um somit den Zustand des empfangenen Symbols festzulegen. Das geschieht für alle Träger parallel. Damit entstehen auf der Empfangsseite wieder die ausgesendeten Bits; möglicherweise mit den Übertragungsfehlern gegenüber dem Sendesignal verfälscht. Da Sender und Empfänger nicht unbedingt auf der gleichen Frequenz sind und sich während einer Aussendung gegeneinander verschieben können, ist auf der Empfangsseite eine Frequenzregelung eingebaut, die einen Frequenzversatz von 200 Hz erlaubt und diesen ausgleichen kann. Die Nachführung ist in den Stufen Grob und Fein aufgeteilt. Der Demapper zerlegt die DQPSK-Symbole der einzelnen Träger in die einzelnen Bits und fügt sie zu einem Datenstrom zusammen.

Um eine schnelle Frequenzabstimmung zwischen Sender und Empfänger zu erzielen, wird auf der mittleren Frequenz des belegten Frequenzbandes

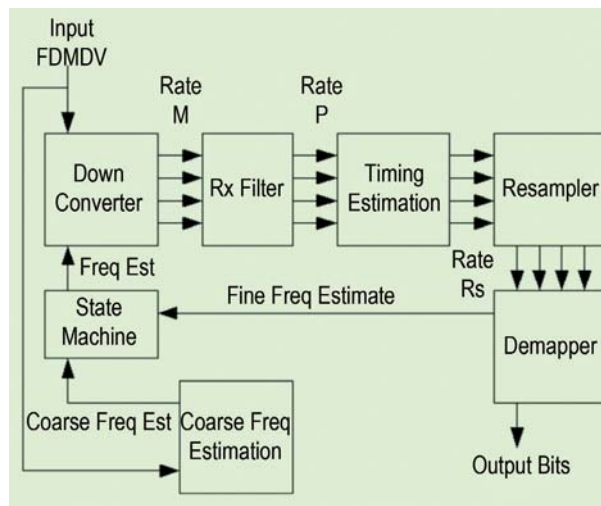


Bild 14: Konzept des FDMDV-Demodem

ein DBPSK-modulierter Pilotträger mit höherer Amplitude ausgesendet. Dieser Kniff führt dazu, auf der Empfangsseite die Übermittlungsbandmitte des übertragenen Signals zu erkennen. Der 10 dB stärkere Träger hat zudem den Vorteil, das für die Aussendung von Mehrträgersignalen wichtige Peak-to-Average-Power-Ratio (PAPR) zu reduzieren. Durch die gleichförmige Modulation mit dem Symboltakt von $f_{symp} = 50 \text{ Hz}$ entstehen zwei Träger bei

$$f_{1,2} = f_M \pm (f_{symp}/2) = (1500 \pm 25) \text{ Hz} \quad (14)$$

die aber bei der Frequenzauflösung der grafischen Empfangsanzeigen zumeist als ein Träger wahrgenommen werden. Der Pilotträger erlaubt beim Empfang eine nahezu augenblickliche Taktrückgewinnung und Nachführung der Frequenz. In dem grafischen FreeDV kann der Benutzer einfach auf diesen Pilotträger klicken, um damit den Empfang mit dieser Mittenfrequenz einzustellen. Die eingestellte Mittenfrequenz des Empfangs wird dabei am unteren Fensterrand mit einem kleinen roten Strich angezeigt.

Keine Synchronisation nötig

Ein wichtiger Vorteil dieses Verfahrens ist, dass es praktisch keine Synchronisationszeiten benötigt. Damit können auch Hörer während eines Sendedurchgangs auf die Frequenz kommen und sofort mithören. Das entspricht der gewohnten Vorgehensweise auf Kurzwelle. Diese für KW wichtige Eigenschaft (ohne Synchronisationszeit) fehlt vielen der zuvor erläuterten Verfahren.

(wird fortgesetzt)



Praktischer Betrieb auf KW

Prof. Dr.-Ing. Michael Hartje, DK5HH

Der abschließende Teil der Beitragsserie widmet sich der Frage, wie man den Sende- und Empfangsbetrieb angeht. Auch für das Finden von QSO-Partnern lesen Sie wertvolle Tipps.

Den Autor erreichen Sie unter Prof. Dr.-Ing. Michael Hartje, DK5HH Nettelbeckstr. 28 28201 Bremen dk5hh@dar.c.de

Im praktischen Betrieb auf Kurzwellen wird das Konzept aus Bild 1 umgesetzt. Dazu stehen zwei Programmsysteme zur Verfügung:

- FreeDV [4] mit einer grafischen Oberfläche fertig compiliert, für fast alle Betriebssysteme (Windows, Mac, mehrere Linux-Distributionen)
- codec2-dev-Ordner enthält alle Sourcecodes für Codec2, FEC, Interleave und Modem. Diese lassen sich ohne grafische Oberfläche mittels Kommandozeile und mit den entsprechenden Aufrufparametern durch Pipelines verbinden. Überdies liefern sie auf Wunsch zusätzliche Informationen über die inneren Verarbeitungsschritte.

Auch FreeDV nutzt die Bibliotheksroutinen des codec2-dev-Ordners, sodass man beide wechselweise und kombiniert benutzen kann. Z.B. Senden per Kommandozeilenskript, Empfangen mit FreeDV und umgekehrt ist möglich.

Beispiel 40-m-QSO mit FreeDV

Zur weiteren Erläuterung sollen jedoch Schirmbilder von FreeDV dienen. Im Folgenden wird dabei nicht auf die Bedienelemente des Programms eingegan-

gen. Das wäre im Rahmen eines eigenständigen Artikels zu behandeln. Der Schwerpunkt soll hier zunächst auf den Empfangsergebnissen und auf den schon mehrfach erwähnten Besonderheiten der Kurzwellenübertragung liegen.

Im **Bild 15** wird der Zeitverlauf des Empfangssignals über 25 s von abendlichen 40-m-QSO im Februar 2013 mit England dargestellt. Die Zeitskala zeigt bei 0 s das aktuelle empfangene Signal. Am oberen Bildende wird das vor 30 s empfangene Signal dargestellt; der Wasserfall läuft von unten nach oben. Der Durchgang des Engländers war vor 6 s zu Ende. Daher werden die unteren 6 s (von 0...6) ohne Signal dargestellt. Lediglich das Brummen des Empfängers ist am linken Rand des Fensters zu sehen. Das Bild wurde nachbearbeitet, sodass die auf dem Bildschirm dargestellten schwarzen Flächen hier entfernt worden sind. Dunkle Flächen stehen für sehr geringe Signalstärke, hellrote Flächen stehen für große Signalstärke. Grün ist etwas schwächer, blau noch schwächer. Die Reihenfolge stark bis schwach: hellrot, grün, blau schwarz. Die für den Empfang genutzte Mittenfrequenz von hier 1500 Hz wird am unteren Bildrand mit einem kleinen roten Strich angezeigt. Deutlich sichtbar ist der stärkere Pilotträger in der Mitte des Bildes. Über längere Strecken ist er rot dargestellt, was ihn als stärkeres Signal als die anderen Träger heraushebt.

Links und rechts der Mittenfrequenz kann man jeweils acht vertikale Linien zählen. Das Signal ist daher mit 1600 Bit/s moduliert. Verfolgt man eine der Trägerlinien von oben nach unten, so ist besonders in der rechten Hälfte zu erkennen, dass die Amplituden der Träger mit etwa 10 s Periodendauer stark schwanken. Im Minimum können die Träger sogar in der Darstellung zu schwarzen Löchern werden – hier sind die Amplituden sehr klein geworden. Andererseits können im Maximum die einzelnen Informationsträger sogar rot dargestellt

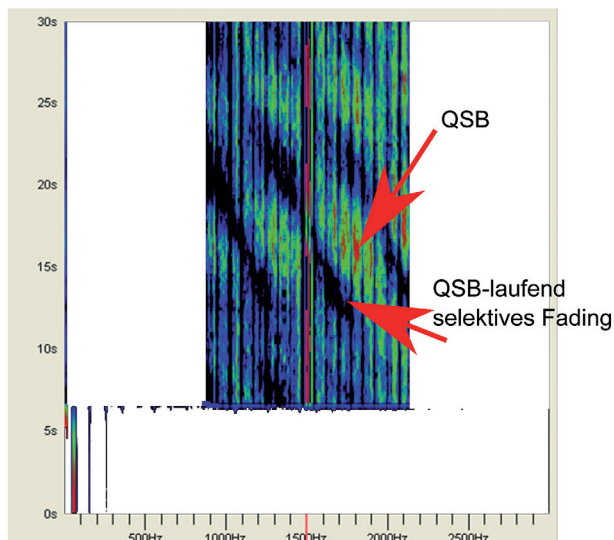
werden, da ihre Amplitude deutlich höher geworden ist. Hierbei handelt es sich um das typische Kurzwellenfading, hier mit einer Periodendauer von etwa 10 s. Vergleicht man die Trägerstärke auf einer Höhe des Bildes, zum Beispiel zum Zeitpunkt $t = 16$ s, so fällt auf, dass der Träger bei 1750 Hz mit sehr starker Amplitude (rot), der Träger bei 1100 Hz aber mit praktisch verschwindender Amplitude (schwarz) empfangen wird. Mit der unteren Frequenzskala ist abzulesen, dass das frequenzselektive Fading hier über der Frequenz binnen 400 Hz einen vollständigen Periodendurchlauf (Minimum, Maximum, Minimum) erreicht.

Das bedeutet, dass über einem Frequenzbereich von 1200 Hz drei vollständige Perioden der Amplitudenschwankungen entstehen; also drei Maxima, zwei bis drei Minima in dieser Breite von 1200 Hz. Dies ist typisch für Kurzwellen und wird als frequenzselektives Fading bezeichnet. Es entsteht durch die sich ständig verändernde Ionosphäre und die sich dann überlagernden Signale von den unterschiedlichen Ausbreitungswegen zwischen Sender und Empfänger.

Bild 16 zeigt zwei Schirmbilder des FreeDV von zwei unterschiedlichen Durchgängen dieses abendlichen 40-m-QSOs mit England. Die beiden eingezeichneten roten Linien verlaufen parallel zu den jeweils auftretenden QSB-Einbrüchen der Einzelträger. Die Steigung der Linien lässt sich mit der Zeitskala an der linken Seite des Bildes beschreiben. Im linken Teilbild beträgt die Steigung vom linken Träger bis zum ganz rechten Träger etwa 10 s, im rechten Bild sind es nur 3 s.

Die beiden Bilder wurden von zwei unterschiedlichen, aufeinander folgenden Durchgängen der englischen Station aufgenommen. Mit den eingezeichneten Linien kann das Fading nun hinsichtlich der Frequenzänderung über der Zeit bestimmt werden. Es ändert sich mit einem ΔF von 120 Hz/s im linken bzw. 300 Hz/s im rechten Bild. Bemerkenswert ist auch, dass über ein Sprachdatenframe – das wäre eine horizontale Linie im linken Bild – teilweise bis zu sieben Träger mit einem „schwarzen Loch“ (z.B. bei $t = 17$ s) eine gleichzeitig schwache Amplitude anzeigen. Im rechten Teilbild ist die

Bild 15: FreeDV-Darstellung-40 m-QSO-Durchgang

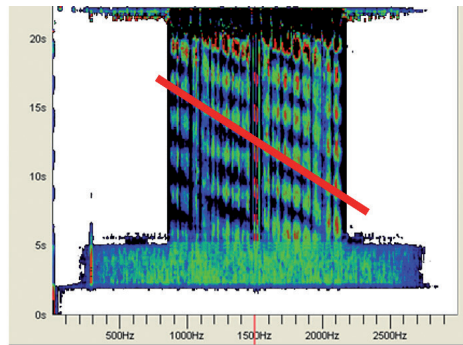


Änderung des Fadings deutlich schneller, sodass weniger Bits eines Sprachframes gleichzeitig betroffen werden. Da 50 Symbole/s mit 25 Sprachdaten-Frames/s ausgesendet werden, sind bei diesem Fading auch mehrere aufeinander folgende Frames betroffen. Dies erschwert die Auslegung von Fehlerkorrekturverfahren erheblich.

Das rechte Teilbild zeigt bis etwa $t = 15$ s neben dem QSB am rechten Rand „blaue Wolken“. Diese zeigen eine splatternde SSB-Station neben der Frequenz an. Auch solche Splatter können zu fehlerhaft übertragenen Informationen führen. Die Decodierung der Sprachdaten unter diesen Bedingungen ist eine anspruchsvolle Aufgabe und geht weit über die relativ ungestörte digitale Übertragung über eine Leitung oder das Fading einer VHF-Verbindung hinaus. Bei der digitalen Sprachübertragung auf KW muss daher dieses Fading berücksichtigt und die fehlerhaft decodierten Bits korrigiert werden.

Durch Verwürfeln der Bits (scramble) innerhalb eines 40 ms dauernden Sprachdaten-Frames (über zwei 32-Bit-wertige Symbole) werden dadurch sehr unterschiedliche Bits von einem „schwarzen Loch“ betroffen. Die fehlergesicherten, wichtigen Bits können dann möglicherweise noch korrigiert werden. Bei den nicht Fehler gesicherten Bits entstehen in der Interpretation des Quellendecodierers (Codec2-Decoder) im Rx dann entweder Artefakte oder Fehler, die man als solche auch hören kann. Der Algorithmus des Codec2 wurde grundsätzlich jedoch so ausgelegt, dass diese nahezu unvermeidlichen Fehler möglichst geringe Auswirkungen auf das Hörergebnis haben.

Das Programm FreeDV zeigt beim Empfang die mittlere Fehlerrate (Bit-Error-Rate BER) an. Bei der Datenübertragung über Drahtleitungen erwartet man eine $BER < 10^{-5}$. Das ist bei einer Funkübertragung jedoch nicht zu erwarten. Es zeigt sich, dass oberhalb einer mittleren Fehlerrate von 0,1 der Empfang der digitalen Sprachdaten stark gestört ist – es entstehen viele Artefakte. Mit einer $BER < 0,03$ – das entspricht 3 % – wird ein praktisch störungsfreier Empfang möglich. Dies erlaubt dann einen für Kurzweile außergewöhnlich störungsfreien Empfang – keine Nebengeräusche, kein Rauschen, obwohl der Signalstörabstand dafür nicht besonders hoch sein muss. In der Modulationsart SSB ist das der typische Empfang eines mittelschwachen



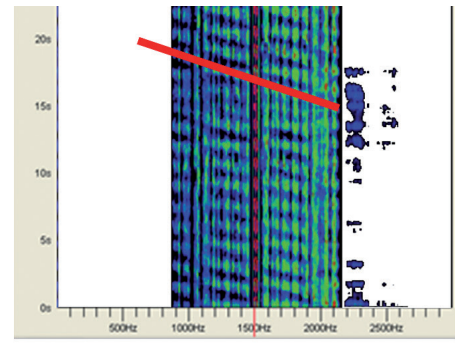
$\Delta F = 120 \text{ Hz/s}$

Signals mit etwa 10 bis 15 dB Signal-/Störabstand, bei dem man meist im SSB-Betrieb schon etwas genauer hinhören muss. Für die Klönrunde wäre dieses Signal wohl eher als „zu schwach“ für eine entspannte Gesprächsrunde beurteilt worden. Immerhin: Mit digitaler Sprachübertragung wird der Empfang praktisch störungsfrei und damit eben doch entspannt möglich.

Verbinden der Codec2-Programme

Als zweite Möglichkeit für die Verwendung von Codec2 soll der Aufruf über die Kommandozeile erläutert werden. Alle Programme des compilierten Ordners `codec2-dev` sind so ausgelegt, dass sie anstelle von Dateien auch die Standard-ein- und -ausgabe des Betriebssystems beim Aufruf mit ansprechen können. Dabei werden die Daten nur sehr kurz gepuffert. Die im Beispielskript (**Kasten**) verwendeten Programme sind `c2enc` sowie `f2mdv`. Die Programme `sox`, `arecord`, `aplay` stammen von anderen Programmpaketen und sollen hier nicht tiefer erläutert werden, da dazu ausführliche Informationen mit „man programmname“ abgerufen werden können. Als Besonderheit sei `sox - swiss army knife of sound exchange` – erwähnt. Es ist ein äußerst mächtiges und sehr vielseitiges Programmpaket für die Sounddatenbearbeitung.

Das hier entwickelte Aufruf-Listing wird als Datei „`mic-c2-1400-mod.sh`“ abgespeichert und so als Shell-Skript aufgerufen. Es stellt damit den Sendezweig zur Verfügung. Das Skript soll kurz erläutert werden. Die einzelnen Programme werden durch das Shell-Skript in der Reihenfolge des Bildes 1 und 12 aufgerufen (verbunden) und übergeben die bearbeiteten Signale über die Pipeline an das folgend genannte Programm. Im Listing ist das Zeichen „|“ das Pipelinezeichen, welches die Programme als Kette aneinander anschließt. Von der Aufnahme der externen Soundkarte (`hw:1`) werden die Abtastwerte auf 8000 Hz mit „`r 8000`“



$\Delta F = 300 \text{ Hz/s}$

mit Bandbegrenzung bei 3,8 kHz gewandelt. Der Aufruf von „`c2enc 1400 - - | \`“ wählt die Eingangsdaten von der Standardeingabe „`-`“ an und schreibt die codierten Daten mit 1400 Bit/s auf die Standardausgabe „`-`“. Die Ausgabe wird auf einer Pipeline an das folgende Programm weitergereicht. Um anzuzeigen das die folgende Zeile noch zum Kommando dazu gehört, wird das Zeichen „`\`“ am Zeilenende erforderlich.

Die vorausschauende Fehlerkorrektur wird hier nicht eingesetzt, sodass direkt hinter dem Codec2 das Modem angeschlossen wird. Der Aufruf „`f2mdv - - 14 | \`“ parametrisiert das Modem in der Pipeline mit 14 Trägern. Auf das Modem folgt die Ausgabe der Modemdaten-Tonfolge. Das Modem erzeugt die Abtastrate von 8000 Samples/s, die Soundkarte erhält hier 44100 Samples/s auf der internen Soundkarte. Die Abtastatenkonversion wird von „`aplay`“ mit den Aufrufparametern vorgenommen. Dieser Ausgang kann bereits an den Mikrofoneingang des Senders angeschlossen werden.

Wenn wir eine Kontrolle des erzeugten Signals vornehmen wollen, können wir nun mit einem zweiten Rechner mit FreeDV die richtigen Soundkarten-Ein- und -ausgänge verbinden und im 1400-Bit/s-Modus die Ergebnisse der digitalen Übertragung im FreeDV-Rechner prüfen. Eventuell werden wir auf dem Rechner, auf dem mit der Kommandozeile die Programme aufgerufen werden, noch mit einem zweiten Terminalfenster das Programm „`amixer`“ aufrufen und dort die richtige Lautstärke einstellen wollen. Sollte in unserem Sender der Frequenzgang stark verzerrt sein, können wir mit einem zusätzlichen Aufruf von „`sox`“ im Anschluss an das Modem einen Equalizer parametrisieren, um die ungleichen Übertragungseigenschaften der NF-Frequenzen im SSB-Sender wieder auszugleichen – näheres siehe „man `sox`“ und dort „`equalizer`“. In gleicher Weise ist es auch möglich den Empfang als Skript aufzusetzen und diese Skriptdatei bei Empfang aufzurufen, wenn vorher der Sendezweig beendet wird.

Bild 16: QSB-Unterschied zwischen 2 Durchgängen auf 40 m

Sendebetrieb mit FreeDV

Die Aussendung eines Frequenzbandes im Durchlassbereich eines SSB-Senders macht es erforderlich dessen Filterdurchlassseigenschaften zu kennen und als „flach“ – also als gleichmäßig – einzuordnen. Das gilt natürlich auch für den Rx-Pfad. Mehrtonaussendung führen aber auch zu deutlichen Unterschieden zwischen dem Mittelwert der Sendeleistung und dem CW-Einton-Wert.

Jeder SSB-Operator kennt, dass SSB-Modulation nicht nur das richtige Mikrofon und dessen Anpassung an den Sender, sondern vor allem die richtige Aussteuerung des Senders erforderlich macht: Zum einen soll keine Übersteuerung, zum anderen aber auch die mögliche Ausgangsleistung des Senders genutzt werden. Wer RTTY oder PSK31 nutzt, hat schon Erfahrungen, wie weit der Sender an- und ausgesteuert werden muss. Ihr Vorteil ist, dass hier immer nur ein Ton ausgesendet wird. Der Sender kann also dafür maximal ausgesteuert werden, eine Übersteuerung und das Ansprechen der ALC des Senders ist im gewissen Bereich unkritisch für das Übertragungsergebnis.

Mehrtonverfahren ähneln sehr der Sprache bei SSB-Betrieb – das müssen wir uns also näher ansehen.

Verzerrungen auf der Sendeseite sind grundsätzlich vermeidbar oder weitestgehend minimierbar. Die richtige Aussteuerung wird damit für eine fehlerfreie Übertragung zwischen Modem (Mikrofoneingang des Senders) und der Sendantenne eine wichtige Voraussetzung für den Sendebetrieb mit digitalen Mehrtonmodulationsverfahren. Verzerrungen im Sender erzeugen Fehler im digitalen Datenstrom, die auf der Empfangsseite ggf. nicht wieder ausgeglichen werden können – zumal sie eben auf der Sendeseite auch vermeidbar sind! Hier wollen wir ansetzen.

Mehrtonaussendungen und deren Pegel

Wir nehmen zunächst einen einfachen Fall an. Es entsteht eine Überlagerung von Sinussignalen. Wir haben ja schon in den vorherigen Teilen der Reihe gesehen, dass die menschliche Sprache ebenfalls eine Zusammenstellung von vielen Mehrtönen (Grund- und Oberschwingungen) ist. Beim SSB-Betrieb gibt es eine Übersteuerungsregelung ALC, die eine ständige Übersteuerung vermindern soll – der Kompressor des kleinen Mannes. Wir wollen zunächst einen Sinuston vom

1500 Hz als Pilotton bei der Übertragung mit FreeDV und dem dort implementierten Modem annehmen. Wenn der Pilotton U_1 etwa 10 dB stärker ist als die benachbarten Informationsträger U_2 , dann entspricht das nach Umrechnung der in dB angegebenen Spannungswerte einem Verhältnis von

$$U_1/U_2 = 10^{10/20} = 3,16 \quad (15)$$

Das bedeutet, dass der Pilotton etwa 3,16-fache Spannung der Nebentöne hat, die benachbarten Informationsträger also nur die 0,316-fache Amplitude des Pilottones.

Das folgende Beispiel soll die Verhältnisse bei der Verwendung von FreeDV verdeutlichen. Der Pilotton habe die angenommene Spannungsamplitude $\hat{U} = 1$. Der Effektivwert der Spannung ist daher $U_{\text{eff}} = 1/\sqrt{2} = 0,7$. Der Kehrwert, das Verhältnis von Spitzenwert zu Effektivwert, hier 1,4, wird als Crestfaktor bezeichnet. Mit einer geringeren Effektivwertspannung (0,7) wird aber auch der Effektivwert des Stromes nur 0,7, sodass die Leistung einen Effektivwert von $P_{\text{eff}} = 1/\sqrt{2} \cdot 1/\sqrt{2} = 1/2$ aufweist. Die Spitzenleistung des Pilottons erreicht also die doppelte Effektivwertleistung. Die Leistung der einzelnen Informationsträger ist 10 dB geringer, also $P_{\text{eff,IT}} = 0,05$. Bei einer Mehrtonaussendung des Pilottons zusammen mit den Informationsträgern werden diese additiv überlagert. Da die Töne nicht alle die gleiche Frequenz aufweisen, wird die gegenseitige Phasenverschiebung zueinander irgendwann aufgehoben. Zu diesem Zeitpunkt summieren sich alle Spannungen. Das führt dann zu einer maximalen Spannung von (Formel 16):

$$U_{\text{max}} = \hat{U} + \sum_{n=1}^{16} U_{IT} = 1 + 16 \cdot 0,316 = 6,06$$

Das **Bild 17** zeigt einzelne sich wiederholende Spitzenwerte mit der maximalen Amplitude von 5,8. Dieser Wert erreicht aus Gründen der Darstellung und Digitalisierung nicht ganz den erwarteten Wert U_{max} . Die Wiederholung alle 14 ms entsteht durch den gleichmäßigen Frequenzabstand aller Töne mit 75 Hz. Deutlich sichtbar ist jedoch auch der Unterschied zwischen dem mittleren Amplitudenwert und dem Spitzenwert. Dies hat zur Folge, dass der Crestfaktor deutlich höher wird als bei einem Sinus-signal. Wenn nun bei der Aussendung eines solchen Signals diese hohen Spitzen komprimiert oder diese gar abgeschnitten werden, entstehen die vorher benannten digitalen Übertragungsfehler

des Senders. Diese Überlagerung von mehreren Tönen (Grundschwingung und Oberschwingungen) haben wir auch bei Sprache beobachtet und in CQ DL 2/14, S. 9ff. ausführlicher erörtert. Daher ist die Anforderung nach linearer (unkomprimierter) Aussteuerung des Senders bereits vom SSB-Betrieb her bekannt. Verzerrung erzeugen Splatter, komprimieren aber auch die Modulation.

SSB-Sender weisen in der Regel eine ALC (Automatic-Level-Control) auf, die beim Ansprechen durch Übersteuerung die Verstärkung der vorgelagerten Senderstufen reduziert. Dadurch kann eine Übersteuerung des Senders vermindert werden. Die ALC ist für den Sprachbetrieb ausgelegt. Bei digitalen Mehrtonverfahren muss diese Übersteuerung jedoch möglichst durch korrekte Einstellung aller Senderpegel vermieden werden. Bereits das Ansprechen der ALC führt dazu, dass Verzerrungen entstehen und damit einzelne Bits nicht mehr korrekt ausgesendet werden.

Das heißt, wir stellen den mittleren Pegel der Sendeleistung bei FreeDV auf einen geringen Wert ein, sodass die Spitzenleistung unverzerrt ausgesendet werden kann. Die Effektivwertleistung wird bei Pilotton mit Informationsträgern mittels quadratischer Addition ermittelt gegenüber 0,5 bei nur einem Ton (Formel 17):

$$U_{\text{eff,tot}} = \frac{1}{2} \sqrt{U_{Pt}^2 + \sum_{n=1}^{16} U_{IT}^2} = \frac{1}{2} \sqrt{1 + 16 \cdot 0,1} = 0,8$$

Es gibt bei digitalen Modulationsverfahren für Mehrtonsignale eine Vielzahl von Verfahren, die Sendeleistung möglichst gut auszunutzen. Alle versuchen, den großen Unterschied zwischen Spitzenwert und dem mittleren Wert zu vermindern. Eine sehr erprobte Methode ist es, zusätzliche Übertragungsdaten so einzufügen, dass die Überlagerung im Spitzenwert vermindert wird. Eine zweite, etwas einfachere Methode wird bei der Übertragung mittels FreeDV genutzt. Es wird eine geringe Fehlerrate bei der Aussendung durch den Sender in Kauf genommen. Ziel ist es dabei die Übertragungsverzerrungen insgesamt niedrig zu halten. Bei FreeDV liegt das Verhältnis von Spitzen- zu Durchschnittsleistung ideal bei mehr als 12 dB. Das würde bedeuten, dass ein Sender, mit einer Ausgangsleistung von 100 W für eine verzerrungsfreie Übertragung von FreeDV nur 6,3 W Durchschnittsleistung abstrahlen dürfte. Wenn wir uns für die

Übertragung nun eine geringe senderseitige Verzerrung erlauben, dann kann das Verhältnis zwischen Spitzen- und Durchschnittsleistung auf etwa 8 dB eingestellt werden. Kappen der Spitzen erlaubt. Untersuchungen von VK5DGR und G3PLX für FreeDV haben gezeigt, dass das möglich ist. Es bedeutet dann: Der 100-W-Tx gibt eine Durchschnittsleistung von etwa 16 W ab. Die dabei entstehenden Bit-Fehlerraten sind vom Kanal und vom Signal-Rauschabstand vom Tx zum Rx inklusive aller Fehler in der Größenordnung von 1 bis 2 %. Dies führt zu praktisch störungsfreier Audioübertragung, dank der Fehlersicherung. Mit dieser Einstellung wird aber die Durchschnittsleistung gegenüber dem linearen Idealfall nahezu verdreifacht, sodass im Mittel eine bessere Übertragungsleistung für die gesamte Aussendung zustande kommt.

Prüfung des Senders

Bei vielen SSB-Tx wird das SSB-Signal nach der Filtermethode mit einem schmalbandigen Quarzfilter erzeugt. Da dieses Filter nicht unbedingt einen idealen Amplitudenverlauf aufweist, wird es nun erforderlich, zumindest einmalig sich Kenntnis darüber zu verschaffen. Dazu wird die Durchlasskurve im Bereich 800...2200 Hz mit einem Tongenerator abgeschätzt.

Am einfachsten ist eine Audiodatei im WAV-Format zu verwenden und diese mit dem Linux-Kommando „play“ auf den Soundausgang Richtung Mikrofoneingang des Tx zu bringen. Alternativ kann man mittels des Linux-Kommandos „play“ einen Sweep (Durchlauf über das Frequenzgebiet) direkt ohne Zwischendatei erzeugen:

```
play -n synth 10 sine 800:2200 vol 0.2
```

Obige Zeile erzeugt einen Sinuston, der innerhalb von 10 s von 800 bis 2200 Hz linear hochläuft. Die Lautstärke wird auf 0,2 herabgesetzt um mögliche Verzerrungen in der Soundkarte zu vermeiden. Die Parameter werden im Einzelnen über „man sox“ erläutert. Mit der mittels „vol 0.2“ gewählten Amplitude bleiben wir nun deutlich unterhalb der Vollaussteuerung des Senders. Beim Durchlaufen wird mit einem geeigneten Leistungsmesser oder an der Stromaufnahme der Endstufe die Anzeige beobachtet. Sie sollte möglichst nicht schwanken. Eventuell muss die Durchlaufzeit auf 30 s erhöht werden. Schwankungen der Leistung auf die Hälfte des beim Durchlauf entstehenden Höchstwertes – das entspricht 3 dB

– sind gerade noch zu vertreten. Anzustreben ist natürlich 0 dB Schwankung! Durch Veränderung der Parameter können wir die Frequenzen der Amplitudeneinbrüche ermitteln und Gegenmaßnahmen planen. Z.B. kann ein linearer Anstieg vielleicht mit einem Tiefpass ausgeglichen werden. Oder wir verwenden eine Zwischenstufe mit sox Parameter-Equalizer. Einzelne schmale Einbrüche im Durchlassverhalten werden wir möglicherweise tolerieren.

Einstellungen des Tx-Zweiges und des Senders

Es ist immer von Vorteil, Sender und Rechner-Soundkarte galvanisch zu trennen. Ob das notwendig ist, muss man individuell prüfen. Für den Amateurfunkbetrieb vorbereitete kommerzielle Interfaces erleichtern dies. Die Verschaltung des Trx mit dem Headset (Mikrofon und Kopfhörer) über eine USB-Soundkarte ist im Bild 1 (CQ DL 2/14) erläutert. In Richtung des Senders mit der internen Soundkarte kann in Richtung Tx dann noch das Interface hinzukommen.

Für die Beschreibung zur Einstellung des Trx soll hier auf der Webseite freedv.org [4] gelinkten Dokumente verwiesen werden. Selbst YouTube-Videos sind dort verfügbar. Wenn der Empfang schon klappt, muss der Tx mit PTT-Steuerung und den richtigen Pegeln für das Mikrofon im FreeDV eingestellt werden. Ein veränderlicher Equalizer im FreeDV erlaubt auch die Betonung der die Sprachverständlichkeit fördernden Frequenzen zwischen 1,5...2,5 kHz. Im Tx schalten wir den möglicherweise vorhandenen Signalkompressor aus – das ist sehr wichtig, wie wir gerade gelernt haben.

Nun muss noch der mittlere Sendepiegel mit der Mikrofonempfindlichkeit des Tx so eingestellt werden, wie vorher beschrieben: Also etwa 1/6 der maximalen CW-Ausgangsleistung bei FreeDV-Sendebetrieb.

Es sei darauf aufmerksam gemacht, dass es einzelne Hinweise zum FreeDV-Betrieb gibt, die eine mittlere Ausgangsleistung bis auf etwa 1/5 oder 1/4 der maximalen Ausgangsleistung vorschlagen. Das sollte jeder selbst erproben. Auf keinen Fall darf nun schon die ALC ansprechen – das wäre zu viel des Guten! Für einen ersten Versuch kann man sicherlich auch auf die programmgesteuerte PTT verzichten. Wer jedoch ein über Hamlib steuerbares Interface oder Trx verfügt, kann die PTT auf einfache Weise von FreeDV aus ansteuern. Wir

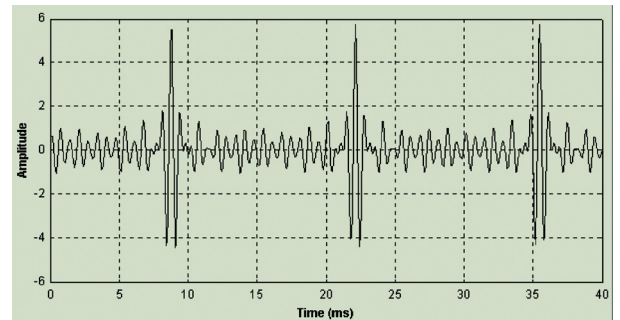


Bild 17: Amplitudendarstellung der Spannung von Pilotton mit Amplitude 1 und 16 Informationsträgern mit jeweils Amplitude 0,316

können versuchen, uns selbst über einen der vielen WebSDR-Empfänger zurückzuhören und zunächst den Empfang mit dem Rechner nur aufzuzeichnen. Das aufgezeichnete Signal können wir später aufrufen und mit FreeDV decodieren. Damit können wir unsere Tx-Einstellungen prüfen.

QSO-Partner finden

QSO- und Testpartner findet man auf der Webseite <http://qso.k7ve.org>. Vorschläge für die derzeit hauptsächlich genutzten Frequenzen werden nach Anmeldung auf der Webseite ausgewiesen. Bei den derzeitigen KW-Bedingungen sollte ein QSO mit USA auf 20 oder 15 m mit 100 W gut möglich sein. Über den Chat auf der Webseite kann man sogar versuchen, einen Testpartner zu überreden, der lediglich sendet – wir selbst empfangen dann nur.

Mit einer größeren Verbreitung dieses kostenlos nutzbaren digitalen Sprachübertragungsverfahrens wird eine Nutzung der Webseite sich allerdings erübrigen. Bis dahin ist das allerdings ein sehr hilfreicher Dienst. Da die experimentelle Erprobung der digitalen Sprachübertragung auf Kurzwellen, wie hier dargelegt, eine sehr niedrige Eintrittsschwelle hat, ist eine wichtige Voraussetzung für eine weitere zügige Verbreitung gegeben. Weitere Entwicklungen, sowohl an FreeDV und an den Verfahren zur Codierung sind möglich und lassen sich leicht von jedem nutzen und auf den neuesten Stand bringen – ein wesentlicher Vorteil der hier beschriebenen Verfahren. Es liegt in der Hand aller Funkamateure dieses Verfahren selbst zu erproben oder auch weiter zu entwickeln. **CQDL**

Listing

```
#!/bin/bash
# TX-Script Mikrofon an ext. Soundkarte, TX an interne SK
echo „Mikrofon mit c2enc 1400 Bit auf Modem“
RATE=44100
arecord -Dplughw:1 -r $RATE -c1 -f S16_LE -t raw - | \
sox -r $RATE -b 16 -e signed-integer -t raw - \
-r8000 -b 16 -e signed-integer -t raw - | \
c2enc 1400 - - | \
fdmdv_mod - - 14 | \
aplay -Dplughw:0 -rate $RATE -t raw -r 8000 -f S16_LE -c 1 -
```