# RADE FOR LAND MOBILE RADIO: A NEURAL CODEC FOR TRANSMISSION OF SPEECH OVER BASEBAND FM RADIO CHANNELS

David Rowe \*

Rowetel P/L Adelaide, South Australia

# Tibor Bece

# Tibicom P/L Goulburn, New South Wales

#### **ABSTRACT**

In the 1990s Land Mobile Radio (LMR) systems evolved from analog frequency modulation (FM) to standardised digital systems. Both digital and analog FM systems now co-exist in various services and exhibit similar speech quality. The architecture of many digital radios retains the analog FM modulator and demodulator from legacy analog radios, but driven by a multi-level digital pulse train rather than an analog voice signal. We denote this architecture baseband FM (BBFM). In this paper we describe a modern machine learning approach that uses an autoencoder to send high quality, 8 kHz bandwidth speech over the BBFM channel. The speech quality is shown to be superior to analog FM over simulated LMR channels in the presence of fading, and a demonstration of the system running over commodity UHF radios is presented.

#### 1. INTRODUCTION

Many land mobile radio (LMR) systems [1] for push to talk (PTT) VHF/UHF voice are built on radio hardware employing the baseband frequency modulation (BBFM) architecture. Applications for LMR include public safety (police, fire, ambulance services), commercial (taxis, mining industry, transport), and personal (off road recreational vehicles, amateur radio). Analog frequency modulation (FM) remains popular (e.g. millions of radios in Australia alone) because of its speech quality, low cost, interoperability and ability to operate without infrastructure. Key requirements for LMR are speech quality, spectral efficiency, and robustness to fading due to vehicle or hand held radio movement. Weak signal/low SNR performance is a secondary requirement.

The ubiquitous BBFM radio architecture still enjoys a great deal of popularity over more modern SDR designs for several reasons:

- 1. It is inexpensive, a complete handset may retail for \$USD40.
- Insensitive to frequency and phase offsets, frequency accuracy, and modulation index. Low cost frequency references can be used.
- Transmitter compliance with spurious emissions requirements is easier than SDRs as there is no carrier feed through, or DAC and mixer non-linearities.
- Only coarse (frame sync) and fine timing estimation is required no phase or frequency estimation, making demodulator implementation straight forward and robust.
- 5. The waveform can be also generated and received by modern IQ SDR radios, so offers a degree of future proofing.

The speech quality of digital LMR systems employing the BBFM architecture is close to analog FM [2][3]; has not evolved since the mid 1990s; and is constrained to an audio bandwidth of less than 4 kHz. Alternatives for higher speech quality include enhancements to LTE 4G communications to support public safety applications [4] such as mission critical PTT (MCPTT). However these systems require handsets with a parallel, non BBFM radio architecture, and cannot operate in remote areas without LTE coverage, or in simplex (handset to handset) mode.

This paper describes a radio autoencoder [5] derived from the RDO-VAE structure of DRED [6]. It sends 8 kHz bandwidth speech over the same BBFM architecture as existing analog and digital LMR systems, but with significant improved speech quality and robustness to fading. Borrowing from our HF work [7], we denote the machine learning parts of the system RADE (RADio autoEncoder). Other approaches using machine learning (ML) to send speech over radio channels include the very low latency work of [8]. Our work differs in the focus on the BBFM channel for the LMR use case, and the use of a vocoder feature set rather than direct encoding of speech.

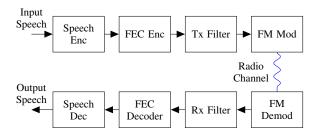


Fig. 1. Classical DSP Speech over BBFM System

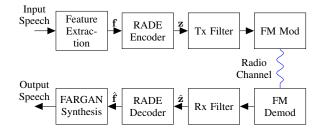


Fig. 2. Radio Autoencoder over BBFM System

Consider the land mobile radio (LMR) system based on classical

<sup>\*</sup>Supported by a grant from Amateur Radio Digital Communications

DSP shown in Figure 1. Speech with a bandwidth of around 3 kHz is sampled at 8 kHz and compressed to a low bit rate using a speech encoder. Forward error correction (FEC) adds redundant bits to protect the sensitive payload speech bits from channel errors. Framing (not illustrated) concatenates several speech encoder frames and adds frame sync and signalling information such as station ID. Payload bits are mapped to a discretely valued (2 or 4 level) sequence of analog shift keyed (ASK) symbols driving an analog FM modulator. The symbols may be filtered to minimise inter-symbol interference (ISI) and constrain the spectrum of the transmitted signal. The receiver employs a limiter and analog FM demodulator to recover the symbols, which are converted to bits with a slicer. The FEC decoder attempts to correct any bit errors, and the speech decoder converts the signal back into a sampled speech signal for replay over a loud-speaker.

Figure 2 presents a radio autoencoder for the BBFM channel. A feature extractor takes speech sampled at 16 kHz and generates a set of vocoder feature vectors (short term spectrum, pitch, voicing)  ${\bf f}$ . The RADE encoder uses a neural network to directly generate a vector of ASK symbols  ${\bf z}$  from the feature vectors  ${\bf f}$ . These ASK symbols are sent over the BBFM channel to the RADE decoder which transforms them back to vocoder features  ${\bf \hat{f}}$ . We employ the FARGAN vocoder [9] for high quality speech synthesis with an audio bandwidth of 8 kHz.

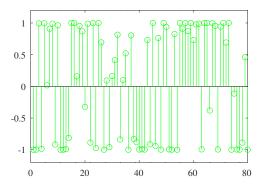
Our contributions are:

- A neural network (radio autoencoder) for sending speech over the BBFM channel that significantly outperforms classical analog and digital alternatives.
- A neural network for sending high quality, 8 kHz bandwidth speech over the BBFM channel, more than double the audio bandwidth of classical analog and digital alternatives.
- 3. An autoencoder that combines the classical DSP functions of quantisation and channel coding to generate discrete time but continuously valued (analog) ASK symbols directly from vocoder features. Unlike classical approaches there is no intermediate bit stream, and the ASK symbols (Figure 3) emerge from the training process rather than being members of a well defined, discrete constellation.
- 4. A linear approximation of the BBFM channel to support training and simulation of ML networks.

The design of the ML components is presented in Section 2. The FM modulation and demodulation process is highly non-linear. In Section 3 we develop a linearised model for ASK pulses over BBFM when the radio signal is subject to AWGN and fading channel impairments. We then describe how this model is used to train the radio autoencoder in Section 4. In Section 5 we use automatic speech recognition (ASR) to evaluate the RADE encoded BBFM over AWGN and fading channels, and compare results to analog FM. In Section 6 we describe a demonstration system running over radio hardware and provide links to speech samples from our system, analog FM, and digital LMR systems for comparison.

#### 2. DESIGN

RADE uses 20-dimensional feature vectors  $\mathbf{f}$  that consist of 18 Bark-scale cepstral coefficients, the pitch period, and a voicing parameter [10]. The RADE encoder and decoder form an autoencoder that is trained to minimise the reconstruction loss  $\mathcal{L}(\mathbf{f}, \hat{\mathbf{f}})$  between the input features  $\hat{\mathbf{f}}$  and the decoded features  $\hat{\mathbf{f}}$ , as defined in Eq. (12)



**Fig. 3.** d=80 ASK symbols from RADE encoder representing 40 ms of encoded speech. The encoder symbols are continuously valued and not part of a discrete constellation.

of [6]. The feature vectors  ${\bf f}$  are generated every 10 ms, concatenated, and passed to the RADE encoder every 40 ms (frame rate of 25 Hz). They are transformed to a d=80 dimensional ASK symbol vector  ${\bf z}$  by a stack of 1D convolutional (conv) and gated recurrent units (GRU), arranged in a DenseNet-like [11] topology. The RADE encoder and decoder are the same as employed for our HF work; further details are provided in [7].

The feature vectors  ${\bf f}$  are updated at 100 Hz, and the dimension d=80 ASK symbol vectors  ${\bf z}$  at 25 Hz, i.e. a symbol rate of  $R_s=2000$  symb/s for the payload speech information. The resources for the ML processing are around 1 Mbyte of read only storage and 32 million multiply-accumulates (MMACs) each for the encoder and decoder, with overall CPU load being dominated by the FARGAN vocoder (300 MMACS) [7].

Our system is not sensitive to the design of the pulse shaping filters, apart from the assumption that they minimise inter-symbol interference (ISI). Not shown in Figure 2 are other components such as the DAC/ADC, RF frequency translation, IF filtering, limiter and power amplifier. All of these components and the pulse shaping filters are assumed to have a unity transfer function with respect to the ASK pulses and minimal impact on system performance. As is common with other digital-over-BBFM systems, no de-emphasis or pre-emphasis is employed. All blocks not marked RADE are implemented with classical analog or digital signal processing, common to existing analog or digital FM radios, and requiring modest amounts of CPU and memory compared to the machine learning (ML) components.

#### 3. BBFM CHANNEL MODEL

In this section we develop a linearised model of the BBFM signal processing in AWGN and multipath fading channels to support simulation and training. We assume an ideal FM receiver that is sensitive only to frequency modulation of the received signal - a reasonable assumption in practice due to the availability of low cost FM demodulators with good performance.

Above threshold, the signal to noise ratio (SNR) at the output of the FM demodulator as a function of the input carrier to noise ratio (CNR) is given by [12]:

$$SNR = 3\beta^2 \overline{x^2} CNR \tag{1}$$

where  $\beta=f_d/f_m$  is the modulation index,  $f_d$  is the maximum deviation,  $f_m$  is the maximum frequency of the modulation signal x(t),  $\overline{x^2}$  is the mean power of x(t), and x(t) is constrained to have a peak of 1 which corresponds to the maximum deviation  $f_d$ . For sinusoidal modulation (common for testing) where  $x(t)=A\cos(\omega_m t)$ ,  $\overline{x^2}=A^2/2$ . Practical FM receiver measurements are often performed using sinusoidal x(t) with A<1, e.g. A=0.6. Equation (1) holds above a threshold SNR which is typically around 12 dB at the FM demodulator output.

The demodulator input CNR is:

$$CNR = \frac{C}{N_0 f_m}$$
 (2)

where C is the carrier power, and  $N_0$  the spectral noise density. Note the noise power is measured in the bandwidth of the modulating signal  $f_m$ . The demodulator output SNR is also measured in a noise bandwidth of  $f_m$ . The somewhat counter-intuitive convention of measuring input CNR in the bandwidth of the modulating signal (rather than the modulated FM signal) allows a simple comparison to linear modulation methods such as AM and SSB [12].

In this paper, we will express SNR as a function of the received signal power R in milliwatts. Expressing the demodulator input CNR as a function of R, and defining the spectral noise density as the thermal noise  $N_0 = kT$  we obtain:

$$CNR = \frac{R}{10^3 kTF f_m}$$
 (3)

where k is the Boltzmann constant, and T=274 K, and F is the noise factor of the radio. Substituting into (1):

$$SNR = \frac{3\beta^2 \overline{x^2} R}{10^3 k T F f_m}$$

$$SNR_{dB} = R_{dBm} + G_{FM}$$

$$G_{FM} = 10 \log_{10} \left( \frac{3\beta^2 \overline{x^2}}{10^3 k T f_m} \right) - NF_{dB}$$
(4)

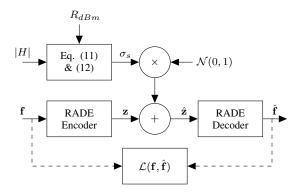
where the noise factor F has been expressed in the more convenient noise figure NF<sub>dB</sub> =  $10\log_{10}(F)$ . Using some typical values for a LMR receiver as NF<sub>dB</sub> = 5 dB,  $f_d = 2.5$  kHz,  $f_m = 3$  kHz, and a sinusoidal modulation signal with A = 1,  $\overline{x^2} = 0.5$ , we obtain:

$$SNR_{dB} = R_{dBm} + 134.41$$
 (5)

For example at the threshold  $SNR_{dB} = 12 \, dB$ ,  $R_{dBm} = -122.41 \, dBm$ .

Our previous experiments show the HF RADE system [7] can provide high quality speech at around 0 dB SNR, suggesting we have ample link margin and should obtain high quality speech over the BBFM channel down to the threshold SNR.

Consider a typical LMR channel from a base station transmitter to a receiver that is pedestrian or vehicle mobile. The propagation path will typically not be line of site, but will reflect off several terrestrial objects (such as buildings) to reach the receiver. The relative phase shifts of the paths will result in multipath fading, which due to the receiver movement will be time varying. The multipath channel typically evolves slowly (e.g. a bandwidth of a few 10's of Hz at 60 km/hr) compared to the symbol rate (kHz), so the rate of change in phase has only a trivial impact on the output of the FM demodulator compared to the data symbols and can be ignored. We can therefore model the multipath channel as Rayleigh distributed magnitude fading |H| and AWGN noise added to the FM carrier. We have chosen



**Fig. 4.** Linearised BBFM model. The fading |H| modulates the noise about the set point defined by  $R_{dBm}$  via Eq. (11) & (12); the magnitude of the symbols  $\mathbf{z}$  is set by the deviation  $f_d$  and remains constant. The loss function  $\mathcal{L}(\mathbf{f}, \hat{\mathbf{f}})$  is used during training to minimise end to end distortion of the vocoder features.

the two path fading model described in the faded channel simulator section of [13]:

$$y(t) = x(t)G_1(t) + x(t - d)G_2(t)$$
(6)

where x(t) is the signal from the transmitter, and y(t) is the output of the multipath fading model.  $G_1$  and  $G_2$  are two band-limited complex Gaussian signals with *Doppler Spread* bandwidth  $B_{ds}$ , and d is the delay spread (path delay) in seconds.  $B_{ds}$  is derived from the vehicle speed as:

$$B_{ds} = 2f_c v/3 \times 10^8 \tag{7}$$

where  $f_c$  is the carrier frequency and v is the vehicle speed in meters per second. By taking the z-transform |H| can be computed as:

$$|H| = |G_1 + e^{-jdF_s}G_2| \tag{8}$$

 $G_1$  and  $G_2$  are time varying and updated at a suitable sample rate.

To simplify simulation and training we model the filtered received symbols from the FM demodulator as the transmitted symbols with additive Gaussian noise:

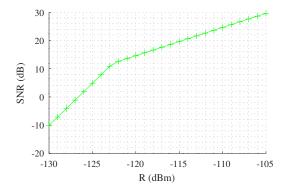
$$\hat{\mathbf{z}} = \mathbf{z} + \mathbf{n} \tag{9}$$

where **n** is a vector of noise samples such that the power of each noise sample is time varying:

$$n_i = \sigma_s^i \mathcal{N}(0, 1) \tag{10}$$

Note this approximates the triangular spectra (noise power increasing with frequency) of FM noise [12, p. 457] with flat noise of the same power. In the following we omit the sample index i for clarity such that  $\sigma_s = \sigma_s^i$ . The mean magnitude and hence power of each symbol in  $\hat{\mathbf{z}}$  at the output of a FM demodulator is a function of the deviation and not the channel CNR. Thus as the CNR is modulated by fading, the mean value of each symbol in  $\hat{\mathbf{z}}$  will be constant, however the FM demodulator output SNR will change, implying that output noise power  $\sigma_s^2$  is modulated as a function of |H| (Figure 4).

Fading may push the input CNR beneath the demodulator threshold. At this point the SNR drops rapidly, and the noise becomes impulsive rather than AWGN. Except for rare deep fades, our system is designed to operate above threshold. For simplicity of



**Fig. 5.** Piecewise SNR versus  $R_{dBm}$  for FM demodulator model in (11) with  $\beta=2.5/3$ , a sinusoidal modulation signal x(t) with  $A=1,NF_{dB}=5$  dB, for an AWGN channel (H=1).

training, we approximate the demodulator SNR using a piecewise linear model where SNR drops rapidly beneath a received power threshold  $T_{dBm}$  corresponding to a demodulator output SNR of 12 dB:

$$SNR_{dB} = \begin{cases} R'_{dBm} + G_{FM}, & R'_{dBm} \ge T_{dBm} \\ 3R'_{dBm} + G_{FM} - 2T_{dBm}, & R'_{dBm} < T_{dBm} \end{cases}$$
(11)

where  $R_{\rm dBm}$  is the mean (set point) received power, and  $H_{\rm dB}=20log_{10}|H|$  is the current magnitude of the fading channel, and  $T_{\rm dBm}=12-G_{\rm FM}.$  In the sub-threshold region (second clause of (11)) a gradient of 3 was selected. As the theoretical performance of the FM demodulator in the sub-threshold region is highly nonlinear and signal dependent, the gradient of 3 was selected as a good approximation to real, measured systems.

Figure 5 plots (11) for an AWGN channel (H = 1).

For simulation of the RADE system over the BBFM channel, we approximate the combined effect of the pulse shaping Tx and Rx filters as simple band limiting to the symbol rate  $R_s$ . Table 1 contains the model parameters we have used for analog FM and RADE simulation over the BBFM channel model.

To calculate the noise power to apply in simulation we solve for  $\sigma_s$ :

$$\frac{S}{\sigma_s^2} = \text{SNR}$$

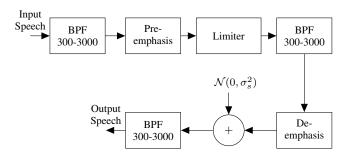
$$\sigma_s = \frac{A}{\sqrt{\text{SNR}}}$$
(12)

where A is the amplitude of the symbol corresponding to maximum deviation  $f_d$ . As |H| evolves over time,  $\sigma_s$  should be re-calculated for every symbol using the latest sample of |H|. The channel simulation procedure is:

- 1. Select a set point (mean)  $R_{\rm dBm}$  and vehicle speed.
- 2. Generate a set of |H| samples at the symbol rate  $R_s$  based on a vehicle speed from the fading channel simulator (8).
- 3. Calculate SNR for each |H| sample using (11).
- 4. Set  $\sigma_s$  for each symbol in the set using (12).

 Table 1. Channel Model parameters

Table 1. Chamber 1.10 der parameters		
Parameter	Analog FM	RADE
Carrier frequency $f_c$	450 MHz	450 MHz
Symbol rate $R_s$	-	2000
Deviation $f_d$	2500	1800
Maximum $x(t)$ freq. $f_m$	3000	2880
Noise figure NF <sub>dB</sub>	5	5



**Fig. 6.** Analog FM simulation. The linearised BBFM model from Eq. (11) & (12) is used to obtain  $\sigma_s$ .

# 5. Apply noise to each symbol using (9).

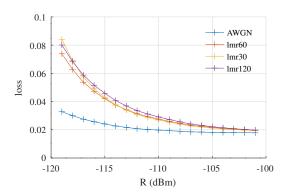
To support comparative testing, analog FM samples can be simulated with the same linearised model as the RADE system, with the noise scaled such that the SNR was referenced to a noise bandwidth  $f_m$  Hz. Speech is band pass filtered and limited to simulate the signal processing in a typical commercial FM radio. Typical peak to average power ratio (PAPR) of the speech tested here is 15 dB before compression and 8 dB after compression.

For a given  $R_{dBm}$ , Eq. (11) provides the demodulator output SNR. To calculate  $G_{fm}$ , we use  $A=1,\overline{x^2}=0.5$ . The SNR given by Eq. (11) has a noise power measured in a bandwidth of  $f_m$ ; the simulated noise is generated at  $F_s=8$  kHz with the same noise density. Speech samples are gain controlled so the peak level just reaches A=1, which results in a measured  $\overline{x^2}=0.07$  for speech signals, around 8 dB less (equivalent to the speech PAPR) than the  $\overline{x^2}=0.5$  for sinusoidal modulation.

#### 4. TRAINING

The linearised BBFM model shown in Figure 4 was used for training the encoder and decoder parameters using the loss function  $L(\mathbf{f}, \hat{\mathbf{f}})$  from [6]. We used a 205 hour speech dataset divided into 4 second sequences for training [6]. For each 4 second sequence, we chose a random set point received signal level uniformly distributed over a 20 dB range  $-100 < R_{dBm} < -120$  to encourage operation at a range of SNRs.

The fading model samples were calculated using a worst case fading simulator path delay  $d=200~\mu s$  and Doppler spreading bandwidth corresponding to a single vehicle velocity of 60 km/hr [13]. We denote this channel lmr60. Despite training at a single simulated velocity, good results (Figure 7) were obtained when the model was tested at a range of simulated velocities between 30 and 120 km/hr, and on AWGN (non fading) channels.



**Fig. 7.** Mean model "loss"  $E[L(\mathbf{f}, \hat{\mathbf{f}})]$  over the 205 hour training set when tested with fading channels based on different simulated vehicle speeds and an AWGN channel. The similar curves for 30, 60 and 120 km/hr indicate the model generalises well.

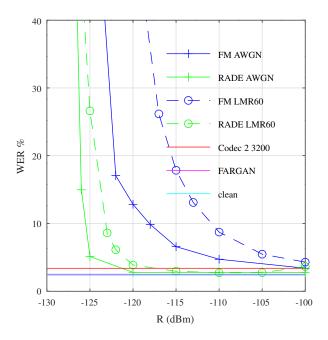


Fig. 8. Word error rate % versus  $R_{\rm dBm}$  for simulated AWGN and lmr60 channels, with clean speech, FARGAN with clean features, and Codec 2 at 3200 bit/s with no bit errors as controls.

#### 5. ASR EVALUATION

This section describes our evaluation of the RADE-BBFM system and comparison to analog FM. We have not included a digital LMR protocol as we had difficulty obtaining a full LMR software stack for simulation. Instead we argue that due to similar speech quality [2][3], analog FM is a reasonable proxy for existing digital LMR systems. We have however included audio samples from a hardware implementation of a digital LMR protocol in our demonstration (Section 6).

Since intelligibility is the primary goal for LMR, we use Automatic Speech Recognition (ASR) to evaluate the performance of the proposed system [7]. Five hundred samples from the Librispeech dataset [14] were passed through RADE and analog FM simulations at a range of received signal levels using the AWGN and lmr60 channels. The linearised channel model derived above was used for these simulations. The output speech was post processed by Whisper ASR system [15], and the Word Error Rate (WER) measured. The Librispeech speech and lmr60 model datasets used for the evaluation were not part of the training dataset. Hilbert compression is applied to the input speech samples for the analog FM simulation, but not the RADE simulation.

Fig. 8 presents the results. Compared to analog FM, RADE is very robust to multipath fading. By drawing lines at constant WER gains of 10 dB can be observed. The WER of RADE approaches the WER of the baseline FARGAN vocoder at high SNRs. The Codec 2 vocoder [16] control simulates the best case performance of classical DSP vocoders employed in current digital LMR protocols. As expected, it matches the WER of analog FM for strong signals.

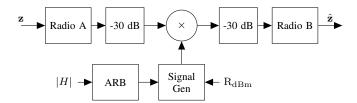
# 6. UHF RADIO DEMONSTRATION

In this section we describe a demonstration of RADE over UHF radio hardware. Our demonstration waveform sends 4800 symbols/s over the BBFM channel (a common symbol rate for digital LMR systems), divided into 40 ms frames, or 192 symbols/frame. The RADE information is sent using 80 symbols/frame, and a 24 symbol unique word is used for frame sync. The remaining symbols are unused in our system but available for ancillary services such as data and station identification. Conventional DSP techniques have been used to perform frame sync and timing estimation.

A pair of commercial off the shelf UHF analog FM LMR radios were configured to provide access to the FM modulator and demodulator. The same radios can be configured to use analog FM to provide a controlled comparison between analog FM and RADE on identical hardware. AWGN noise is generated by physical processes in the radio receiver front end. To provide adequate RF isolation the transmit signal is frequency shifted 20 MHz using a mixer and signal generator. The signal generator drive level to the mixer is used to establish the set point level  $R_{dBm}$ . To implement fading the signal generator is amplitude modulated by an arbitrary waveform generator (ARB) driven by calibrated fading magnitude samples generated using Eq. (8).

RADE, analog FM, and DStar [17] (DStar is a digital LMR system) were tested on AWGN and lmr60 channels at a range of received signal levels. In each instance the received levels were confirmed by a calibrated spectrum analyzer. For each test, a 10 second file containing 4 sentences (2 spoken by male, 2 by a female) was played through the system in real time.

A selection of demonstration audio samples are available online at [18]. Informal listening tests using the demonstration samples



**Fig. 9.** Configuration for UHF radio demonstration. We inject and sample RADE symbols  $\mathbf{z}$  and  $\hat{\mathbf{z}}$  at the analog FM modulator and demodulator.

show RADE outperforming analog FM and the DStar digital LMR system by a large margin.

#### 7. CONCLUSION

In this paper we have described a ML based system for speech communication over the ubiquitous, low cost BBFM radio architecture. Simulations and our demonstration over real radio hardware indicate it outperforms analog FM on AWGN and multipath channels by a significant margin. This provides a "drop in" upgrade path for analog FM and classical DSP digital radios based on BBFM hardware. For the digital radio case the discrete ASK symbols from classical vocoder and FEC components can be replaced by the RADE symbols to provide an upgrade in speech quality and robustness.

The source code for the RADE BBFM system (including channel simulation and model training) is available on GitHub [19] in the dr-bbfm branch. We would like to acknowledge the contribution of George Karan in conceiving and developing this project and supporting the UHF radio demonstration.

The complexities and non-linearities in BBFM systems make exact modelling difficult, so several approximations were made to simply training and simulation. Our simplified linear model does not take into account non-linearities such as the effect of IF filtering on the baseband waveform; we use an approximation of flat versus triangular noise spectra, and a simplified model of the sub-threshold region of FM demodulator operation. The expressions for FM SNR [12] are only tractable for sine wave modulating signals. However the significant improvements we have achieved over analog FM indicate these approximations are acceptable, and more accurate models of the BBFM process may not be justified.

The CPU requirements for RADE are high compared to classical vocoders. As further work investigating optimisations of the RADE encoder, decoder and FARGAN vocoder could significantly reduce this CPU complexity. However the CPU and memory requirements are already a fraction of that available in common personal communications devices such as mobile (cell) phones, and we anticipate will soon be commonplace in modern LMR radios.

Further work includes research into repeater operation and trunking RADE voice traffic over digital networks. This would require quantising the continuously valued ASK symbols to a set of discrete levels. As the waveform is robust to significant AWGN noise we anticipate quantisation noise will not unduly affect performance.

# 8. REFERENCES

[1] Kunagorn Kunavut, "An overview of digital trunked radio: Technologies and standards," *Journal of Industrial Technol-*

- ogy, vol. 10, no. 2, pp. 111-121, 2014.
- [2] David J Atkinson, "Intelligibility of the adaptive multi-rate speech coder in emergency-response environments," Tech. Rep., Institute for Telecommunication Sciences, 2012.
- [3] John M Vanderau, "Delivered audio quality measurements on project 25 land mobile radios," Tech. Rep., Institute for Telecommunication Sciences, 1998.
- [4] Abdallah Jarwan, Ayman Sabbah, Mohamed Ibnkahla, and Omneya Issa, "LTE-based public safety networks: A survey," *IEEE communications surveys & tutorials*, vol. 21, no. 2, pp. 1165–1187, 2019.
- [5] Timothy J. O'Shea and Jakob Hoydis, "An introduction to deep learning for the physical layer," 2017.
- [6] Jean-Marc Valin, Jan Büthe, Ahmed Mustafa, and Michael Klingbeil, "DRED: Deep REDundancy coding of speech using a rate-distortion-optimized variational autoencoder," *IEEE Journal of Selected Topics in Signal Processing*, 2024.
- [7] David Rowe and Jean-Marc Valin, "RADE: A Neural Codec for Transmitting Speech over HF Radio Channels," 2025.
- [8] M. Bokaei, J. Jensen, S. Doclo, and J. Østergaard, "Low-latency deep analog speech transmission using joint source channel coding," *IEEE Journal of Selected Topics in Signal Processing*, 2025.
- [9] Jean-Marc Valin, Ahmed Mustafa, and Jan Büthe, "Very low complexity speech synthesis using framewise autoregressive GAN (FARGAN) with pitch prediction," 2024.
- [10] J.-M. Valin and J. Skoglund, "LPCNet: Improving neural speech synthesis through linear prediction," in *Proc. Interna*tional Conference on Acoustics, Speech and Signal Processing (ICASSP), 2019, pp. 5891–5895.
- [11] Gao Huang, Zhuang Liu, Laurens Van Der Maaten, and Kilian Q Weinberger, "Densely connected convolutional networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 4700–4708.
- [12] Paul B Crilly and A Bruce Carlson, Communication Systems, McGraw-Hill Education, 2009.
- [13] Telecommunications Industry Association et al., "Digital C4FM/CQPSK transceiver measurement methods," ANSI/TIA-102.CAAA-E-2016, 2016.
- [14] Vassil Panayotov, Guoguo Chen, Daniel Povey, and Sanjeev Khudanpur, "Librispeech: an asr corpus based on public domain audio books," in 2015 IEEE international conference on acoustics, speech and signal processing (ICASSP). IEEE, 2015, pp. 5206–5210.
- [15] A. Radford, J. W. Kim, T. Xu, G. Brockman, C. McLeavey, and I. Sutskever, "Robust speech recognition via large-scale weak supervision," 2022.
- [16] D. Rowe, "Codec 2 Algorithm Description," https:// rowetel.com/downloads/codec2\_doc.html.
- [17] The Japan Amateur Radio League Inc, "Digitalization technology standard for amateur radio," 2016, https://www.jarl.org/English/7\_dstar/STD6\_0\_E.pdf.
- [18] T. Bece D. Rowe, "RADE Over Baseband FM," https://freedv.org/rade-bbfm/.
- [19] D. Rowe, "RADAE (RADE) GitHub Repository," https://github.com/drowe67/radae.