

EFFICIENCY IMPROVEMENT OF HF COMMUNICATIONS USING COGNITIVE RADIO PRINCIPLES

L. Melián-Gutiérrez*, S. Zazo[†], J.L. Blanco-Murillo[†], I. Pérez-Álvarez*, A. García-Rodríguez* and B. Pérez-Díaz*

* Instituto para el Desarrollo Tecnológico y la Innovación en Comunicaciones (IDeTIC). Universidad de Las Palmas de G.C. Campus de Tafira, 35017-Las Palmas G.C. (Spain). e-mail: lmelian@idetec.eu

[†] Universidad Politécnica de Madrid (UPM). Dpto. SSR, E.T.S. Ingenieros de Telecomunicación. Ciudad Universitaria, 28040-Madrid (Spain). e-mail: santiago@gaps.ssr.upm.es

Keywords: Cognitive Radio, HF, Hidden Markov Model, Spectrum Activity Model.

Abstract

Cognitive Radio principles can be applied to HF communications to make a more efficient use of the extremely scarce spectrum. In this contribution we focus on analyzing the usage of the available channels done by the legacy users, which are regarded as primary users since they are allowed to transmit without resorting any smart procedure, and consider the possibilities for our stations -over the HFDVL (HF Data+Voice Link) architecture- to participate as secondary users. Our goal is to enhance an efficient use of the HF band by detecting the presence of uncoordinated primary users and avoiding collisions with them while transmitting in different HF channels using our broad-band HF transceiver. A model of the primary user activity dynamics in the HF band is developed in this work. It is based on Hidden Markov Models (HMM) which are a powerful tool for modelling stochastic random processes, and is trained with real measurements from the 14 MHz band.

1 Introduction

Standard HF communications make use of ALE (Automatic Link Establishment) protocol that is frequently presented as an example of a primitive form of cognitive radio. It is based on Listen Before Transmit strategy to access the spectrum and makes use of sensing, probing and monitoring techniques to assess channel utilisation and channel quality. Even if HF stations making use of ALE protocol, there are multiple collisions between them as they only transmit on their assigned channels that are expected to be available according to the experience of the radio operator. Considering the previous, it is likely that as long as cognitive radio principles are introduced in HF stations, the radio frequency spectrum could be exploited in a more efficient way because HF stations will be aware of their surrounding environment and will learn from it. In order to achieve such improvements in the band, these cognitive stations shall have the capability to change their operating parameters in order to adapt their transmissions to the available spectrum holes by using the acquired knowledge.

In cognitive radio systems two types of users are distinguished: primary and secondary users. Primary users are those users that are holders of a license for a particular frequency band and therefore shall have prime access to it. On the contrary, unlicensed users transmitting in that band are taken as secondary users as they are allowed to make use of it, but required not to interfere communications coming from primary stations. In this contribution, due to the trans-horizon behaviour of HF communications and the limited bandwidth where all HF users around the world can transmit, we shall assume that the difference between primary and secondary users relies on the smart and cognitive capabilities of the HF user. Consequently, we will consider the primary ones to be the legacy users which transmit without resorting any smart procedure, and our stations using the HFDVL (HF Data+Voice Link) architecture [3], which are also licensed for certain HF bands, will be henceforth considered as secondary users of the accessible frequency channels. So, our stations must be able to detect in both transmitter and receiver sides when the primary users start and finish their transmissions in order to send our own data packages by filling silence slots within primary transmissions. This will require us to use our broad-band HF transceiver [4] to select those channels predicted as available. By this way, we will enhance the efficient use of the HF band and the performance of our HF communication system.

A model of the primary user dynamics in the band shall be extremely useful to make the best from the acquired knowledge in predicting the activity of the primary user in the channel. An activity model is derived in this work for the HF band based on a set of Hidden Markov Models. It has been trained and validated with real data from the HF band and according to the peculiarities of the particular problem we are considering.

Regarding the modelling and detection problems, Hidden Markov Models (HMM) are widely used for speech recognition [5], and have recently been introduced into cognitive applications for primary user detection and behaviour prediction, where they have achieved remarkable results. Despite this, most of the research in Cognitive Radio is related to communication bands above the HF band. Quite recently Koski and Furman introduced the challenges and

opportunities of applying cognitive radio principles to HF communications in [2]. However, to the best of our knowledge, no proposal of a real or simulated cognitive system for the HF band has been published yet.

2 Hidden Markov Models

HMMs are a powerful and robust tool for modelling stochastic random processes as they are able to model a large variety of processes achieving high accuracy with relatively low model complexity. They have been extensively used in a myriad signal processing applications during the last 20 years, mainly for fitting experimental data onto a parametric model which can be used for real-time pattern recognition, and to make short-term predictions based on the available prior knowledge.

A Hidden Markov Model is defined as a doubly embedded process with an underlying stochastic process that is not observable. These hidden processes (states) can only be evaluated through another set of processes that produce sequences that actually can be observed [5]. Formally, an HMM for discrete symbol observations is defined by the following elements:

- **A**: State transition probability distribution matrix $\mathbf{A} = \{a_{ij}\}$ of size N^2 where N is the number of states in the model and

$$a_{ij} = P[q_{t+1} = S_j | q_t = S_i], \quad 1 \leq i, j \leq N; \quad (1)$$

The set of individual states is $S = \{S_1, S_2, \dots, S_N\}$ and the state at time t is denoted as q_t .

- **B**: The observation symbol probability distribution for state i , $\mathbf{B} = \{b_i(k)\}$, where

$$b_i(k) = P[O_t = V_k | q_t = S_i], \quad 1 \leq i \leq N; 1 \leq k \leq M, \quad (2)$$

O_t is the observation at time t and M is the number of distinct observation symbols per state. The observation symbols correspond to the physical output of the system being modelled and are denoted as $V = \{V_1, V_2, \dots, V_M\}$.

- $\boldsymbol{\pi}$: The initial state probability distribution $\boldsymbol{\pi} = \{\pi_i\}$ where

$$\pi_i = P[q_t = S_i], \quad 1 \leq i \leq N. \quad (3)$$

- **O**: An observed sequence

$$O = O_1 O_2 \dots O_T \quad (4)$$

where each observation O_t is assumed to be a realization of one of the symbols from V , and T is the number of observations in an observed sequence.

Thus, a complete definition of an HMM involves: two model parameters (N and M), the specification of the observation symbols set and the specification of \mathbf{A} , \mathbf{B} and $\boldsymbol{\pi}$. In practise the compact notation used for an HMM will be

$$\lambda = (\mathbf{A}, \mathbf{B}, \boldsymbol{\pi}). \quad (5)$$

Regarding the structure of the transition matrix \mathbf{A} , particular configurations of the HMMs can be identified: ergodic or fully-connected HMM where every state can be reached from any other state after a finite number of steps, and left-right or Bakis HMM which has the property that as time increases the

state index either increases or stays the same. The later was introduced to fit signals which change over time in a successive manner, and is characterised by a transition matrix which forbids non-causal transitions within the models.

Moreover, three major tasks must be fulfilled for a Hidden Markov Model defined as (5) to be useful in practical applications:

1. *Evaluation*: Given an observed data sequence and a set of models to be compared, the probability that the provided sequence was produced by each of them, $P(O | \lambda_k)$, is computed as the cumulative product of the likelihood obtained by each evaluated sample. This is calculated through the sum of the so called, terminal forward variables, $\alpha_T(i)$, defined in the ‘‘forward-backward’’ algorithm [5].
2. *Decoding*: Given an observation sequence, we choose a state sequence that is optimal in some meaningful sense. By doing this, we try to uncover the hidden part of the model according to some criteria. The most widely used criteria is to find the best state sequence that maximises the probability $P(Q, O | \lambda)$ using the well-known Viterbi algorithm [5].
3. *Learning*: Also referred as the training of the model for which we attempt to find the parameters which best describe the observed sequence (the training sequence). An iterative procedure such as the Baum-Welch method [5] is used to locally maximise the likelihood of the model $P(O | \lambda)$.

3 Data acquisition and processing

Real measurements from the HF band have been recorded to train and test the proposed model based on HMMs. We have used our own broad-band transceiver [4], along with a Vector Signal Analyzer (VSA) and SystemVue Software, both from Agilent Technologies, to collect the time data from the antenna and use it to extract spectral information from the HF band by means of the Fast Fourier Transform (FFT). Our broad-band HF transceiver was required in the measurement system to collect data from the antenna and then transmit it to the Vector Signal Analyzer because the sensibility of the VSA was not suitable for our purposes and was not able to detect most of the HF signals in the band.

Sequential measurements of the 14 MHz amateur band were collected during three days. Each measurement was 10 minutes long and between measurements a 15 minutes lag was introduced. The reasons for choosing this particular band are mainly two: the limited bandwidth of our antennas and the high-activity in this band that will be useful for the validation of the proposed model. In order to acquire data from the whole band we selected 14175 kHz as the central frequency and a span of 500 kHz, although the collected span by VSA is wider than 500 kHz. With these parameters we obtained measurements of the amateur band and also channels outside of it which correspond to other radio stations as it can be seen in the example of Figure 1.

The information captured by the VSA was processed to obtain an estimation of the spectrum power of the whole

acquisition band. Further on, both frequency and time domain processing were applied to the collected data to obtain a time-frequency representation of the activity in each channel. Estimates on the mean power of a 3 kHz channel over a two seconds time-window were obtained in this processing step.

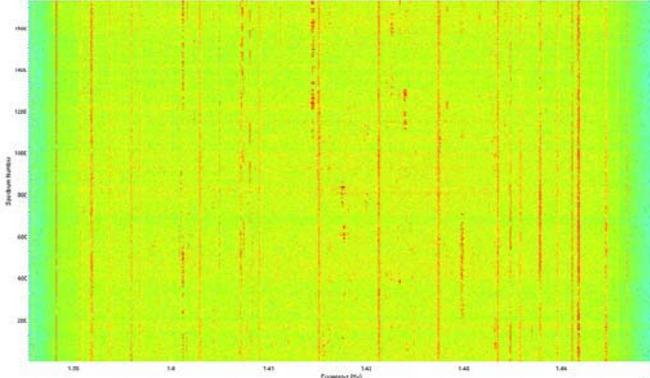


Figure 1: Example of the acquired HF spectrum of the 14 MHz band with central frequency 14175 kHz, span of 500 kHz and a duration of 10 minutes.

Once these estimates are obtained, the signal detection problem is set out to make a decision about the activity in the channel based on averaged power estimates. For this purpose, two hypotheses were defined:

$$H_0: \quad x[n] = w[n] \quad n = 1, \dots, N-1 \quad (6a)$$

$$H_1: \quad x[n] = w[n] + s[n] \quad n = 1, \dots, N-1 \quad (6b)$$

H_0 holds when there is only noise, $w[n]$, in the channel at sample n , while H_1 is true when there is a signal, $s[n]$, in the channel at sample n . The detector that maximises the detection probability in identifying activity on a given channel for a given false alarm probability is defined through the likelihood ratio test formulated in (7) as specified by the Neyman-Pearson lemma [1]

$$L(x) = \frac{p(x; H_1)}{p(x; H_0)} > \gamma \quad (7)$$

where $p(x; H_1)$ is the probability distribution function for occupied channels, $p(x; H_0)$ the probability distribution for only-noise channels and the threshold γ of the detector is derived from the false alarm constraint

$$P_{FA} = \int_{\{x: L(x) > \gamma\}} p(x; H_0) dx = \alpha \quad (8)$$

The probability distributions for both hypotheses, $p(x; H_0)$ and $p(x; H_1)$, were estimated by computing the normalised histogram of the power of only-noise samples and occupied-channel samples, respectively. As a result, threshold γ was defined for each acquired wide-band measurement.

Due to the time restrictions in the acquired measurements, the processed channel sequences have a length of ten and nine minutes. A segmentation of these sequences is proposed here to divide them into observation sequences of one minute in order to increase the number of available data and considering that a minute long observation was long enough to have an accurate representation of the temporal behaviour of primary users in the HF band.

Moreover, in order to reduce the variability of the submodels in the training stage, we also propose a classification of the observation sequences prior to the stochastic modelling of the primary user dynamics. This classification is based on the expected secondary-user behaviour in a cognitive radio system, that is:

- *Available channels* where the secondary user can transmit.
- *Unavailable channels* where a primary user is in the band and the secondary user cannot transmit.
- *Partially available channels* in which there are time intervals during which the secondary user can transmit until a primary user appears.

The criteria followed in the proposed classification is based on the percentage of time that a channel is occupied by a primary user. Two thresholds were defined: one for the maximum percentage of occupation in available channels and another for the minimum percentage of occupation in unavailable channels. In order to derive both thresholds, it is sensible to consider the worst case of channel occupation in HF data communications. The estimation of this worst scenario requires us to assume a predefined value for the maximum length of a data frame. In this work, the frame of maximum length defined by the NATO Standardisation Agreement STANAG 5066 [6] has been used. This standard establishes the profile for professional HF data communications and defines frames, called D-PDU, that have 46 bytes of overhead and up to 1023 bytes of user data. On the other hand, most HF data communications are established at data rates of 600 bps or 1200 bps. With these parameters, the longest possible D-PDU frame would be 14.25 seconds long (1069 bytes at 600 bps) which represent the worst case of channel occupation in our scenario. Therefore, we considered observation sequences with at least 45 seconds occupied within a minute to be unavailable channels as we won't be able to transmit the largest frame during the remaining 15 seconds. Similarly, available channels were defined as those observation sequences with a maximum of 15 seconds occupation in a minute. Finally, the remaining observation sequences were classified as partially available channels, as the secondary user could transmit for less than 45 seconds to avoid collisions with primary users.

4 HF primary user dynamics model

As it has been described, Hidden Markov Models were used in this contribution as a tool to model the primary user activity dynamics in the 14 MHz band. This model should be used by our HFDVL stations, acting as secondary users according to the scenario previously outlined, to predict the presence of a primary user in the band.

Once the data from the HF band has been processed and classified, an interference model can be properly built. According to the proposed classification of the observation sequences, the model was defined as an ergodic HMM with three states interconnecting three underlying submodels, one for each class (i.e. available, unavailable and partially available channel) as it can be seen in Figure 2. Moreover,

each submodel was trained with a set of observation sequences of a minute long from the corresponding class. Consequently, the first submodel is meant to characterise available channel sequences, the second submodel for unavailable channel sequences and the third one for partially available channel sequences. These submodels were implemented as left-right HMM as the particular structure of the transition matrix they exhibit is quite suited to model the time evolution of the samples of the observation sequences, much better than ergodic models can, and with a reduced number of parameters to be trained.

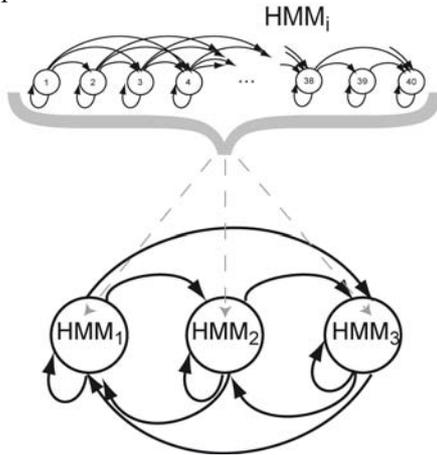


Figure 2: HF primary user dynamics model based on HMMs.

The proposed model for primary user dynamics is built through the combination of a high-level, ergodic HMM with three states, each of them corresponding to available, unavailable or partially available channels respectively, and a set of three left-right submodels. Each state in the high-level model emits a minute long observation sequence which corresponds to the one generated by the submodel for that state. However, to simplify the training and evaluation procedures of the high-level model, it was trained as an independent HMM where each state will just emit one single value representing the state, independently from the scores provided by the low-level submodels. That is, state S_1 only emits symbol '1' and it represents the observation sequence generated by submodel 1. This also happens for states S_2 and S_3 emitting symbols '2' and '3', respectively. So, the observation matrix for the high-level ergodic model, \mathbf{B} , is actually the identity matrix. With this specification the observation symbols generated by the high-level model are also the transitions between its states, that is, the transitions between the underlying submodels.

Furthermore, while submodels were trained to characterise observation sequences of a minute long from available, unavailable or partially available channels, the high-level model was trained to characterise the evolution of a particular channel for ten and nine minutes long sequences, where states are classified according to one minute long sequences.

Once designed, the proposed model was trained and validated as it is described in the following subsections.

4.1 Model training

The learning problem was solved in this first stage in order to find the parameters of the HMM defined as (5) that maximise $P(O|\lambda)$, that is, maximise the probability that the observation sequences used as training sequences were generated by the model. The Baum-Welch algorithm [5], based on the Expectation-Maximisation method, was used to train both submodels and the high-level model. Also, randomly initialised matrices were used for the first iteration of the Baum-Welch method, as no prior knowledge on the structure of the sequences was provided, and the amount of collected data seemed to be fairly enough.

Two different protocols were followed to train the models as the selected structure for the high-level model was different from the submodels structure. As an example, the number of states of the high-level model was prefixed -3- due to the prior classification of the observations, whereas the number of states for each submodel had to be chosen in advance. Therefore, we had to choose the number of states for the HMM submodels prior to their training by looking for a maximum in their likelihood curve. Besides, when considering the training of the high-level model, we shall look for the model corresponding to the maximum likelihood and between different models with matrices initialised to different random seeds. Once the models were trained, their respective likelihood scores were evaluated according to the evaluation task and compared among the different models to select the one with the maximum likelihood value, i.e. the one with highest $P(O|\lambda)$.

The best number of states for the submodels, based on left-right HMM with up to three-states-long transitions was estimated using all observation sequences of one minute long as training sequences. In these left-right HMM submodels, matrices \mathbf{A} and \mathbf{B} were randomly initialised with a common seed, while π was initialised as $(1, 0, \dots, 0)$ due to the fact that these submodels always begin at the first state. Different models with a number of states varying from 20 to 45 were trained with their own observation sequences of a minute long: available channel submodel was trained with all sequences classified as available channels, and the same procedure was followed for unavailable channel and partially available submodels, respectively. Once trained, submodels' likelihoods were evaluated attending to the plots of Figure 3. Unavailable channel submodel and partially available channel submodel likelihoods reach a local maximum when forty states are included, while the likelihood of available channel submodel is very close to zero for any state, i.e. this submodel can perfectly model most of the available channel sequences. Furthermore, if all submodels have the same number of states, these can be easily compared by simply comparing their corresponding estimated likelihood scores. Therefore, all submodels were considered to be left-right Hidden Markov Models including forty states and up to three-states-long transitions as it is represented in Figure 2.

Once the number of states of the submodels had been chosen, each submodel was trained with a 70% of the observation sequences. The Baum-Welch method was applied to randomly initialised HMM matrices **A** and **B** with the same seed used in the previous step. Matrices of the trained submodels were obtained after ten iterations of the algorithm to prevent over-fitting. It has been checked that these matrices represent stable models, that is, there is a transition with a high probability in each row of the transition matrices and if the transition probabilities along the matrix are analysed, the final state is always reached for any observation sequence.

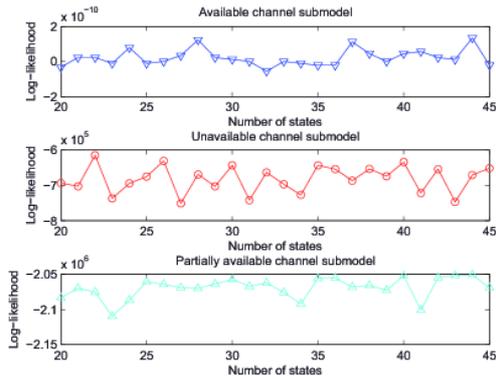


Figure 3: Evolution of submodel log-likelihoods versus the number of states.

4.2 Validating the proposed model

From the initial set of collected data, once the 70% have been put aside for the training of the HF primary user dynamics model, the remaining 30% were left to evaluate the trained model and validate it. As the HF primary user dynamics model is based on three submodels and a high-level model, the evaluation protocol requires two steps: first, submodel likelihoods are evaluated with sequences one minute long and afterwards, the high-level model probabilities are calculated for each observation sequence of ten or nine minutes.

To compare all submodels, the evaluation problem must be solved taking the submodel with the highest likelihood as the local solution. A matrix containing values {1, 2, 3} was built from the results coming from submodels decoding, with its rows corresponding to the observation sequences of nine and ten minutes. Due to the definition of the high-level model, these observation sequences are also equivalent to the transitions of the high-level model. Computing the difference between this generated matrix and the real one, we were able to estimate the percentage of wrong decisions, which actually corresponds to 5% for one minute long test observation sequences.

As previously stated, the resulting matrix will contain the observation sequences for the high-level model and can be used to evaluate it. At this point it is not possible to compare models in order to solve the evaluation problem but, as these observation sequences were the same as the transition sequences due to the initial restriction on matrix **B** of the model, the decoding problem can be evaluated using the

Viterbi algorithm [5]. This algorithm was developed to identify the state sequence which maximises the probability $P(Q, O | \lambda)$, so that this state sequence can be directly compared to the observation matrix. The percentage of errors obtained was actually the same as in the previous step, so, the high-level model can emit sequences such as the ones used in this test. Furthermore, if we think of it as a pseudo-random sequence generator, we could guarantee that it would be able to emit random sequences statistically close to the distribution of the real measurements used for its training.

Conclusions

In this contribution Hidden Markov Models have been used to model primary user activity dynamics in the 14 MHz band which anticipates the adoption of cognitive radio principles in HF communications. This model would be used by our HFDVL stations, acting as secondary users, to predict the presence of a primary user, and consequently, to make the best use from the available frequency channels.

The proposed model is built from three interconnected submodels which describe three types of channels: available channels, unavailable channels and partially available channels. It has been trained and validated with real measurements collected from the 14 MHz amateur band and only 5% test sequences were wrong modelled.

Acknowledgements

This work has been mainly supported by the Spanish Ministry of Science and Innovation project TEC2010-21217-C02 and the Universidad de Las Palmas de Gran Canaria with a scholarship for postgraduate studies. The Spanish Ministry of Science and Innovation in the program CONSOLIDER-INGENIO 2010 under the grant CSD2008-00010 COMONSENS have also partially contributed.

References

- [1] S. M. Kay. *Fundamentals of Statistical Signal Processing: Detection Theory. Vol. II.* Prentice-Hall. (1998)
- [2] E. Koski, W. Furman. "Applying Cognitive Radio Concepts to HF Communications". *11th International Conference on Ionospheric Radio Systems and Techniques.* (2009)
- [3] I. Pérez-Álvarez *et al.* "Real link of a high data rate OFDM modem: description and performance". *11th International Conference on Ionospheric Radio Systems and Techniques.* (2009)
- [4] B. Pérez-Díaz *et al.* "Design and construction of a broadband (1 MHz) digital HF transceiver for multicarrier and multichannel modulations". *11th International Conference on Ionospheric Radio Systems and Techniques.* (2009)
- [5] L. Rabiner, B.-H. Juang. *Fundamentals of Speech Recognition.* Prentice-Hall. (1993)
- [6] STANAG 5066: Profile for HF Radio Communications Ed. 2, North Atlantic Treaty Organization, NATO. (2008)