



UNIVERSIDADE FEDERAL FLUMINENSE
ESCOLA DE ENGENHARIA
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA ELÉTRICA
E DE TELECOMUNICAÇÕES

RAFAEL SOUZA AZEVEDO

**Aplicação de Modelos Ocultos de Markov na
Caracterização Banda Larga do Canal Rádio
Móvel em Diferentes Ambientes**

NITERÓI

2023

UNIVERSIDADE FEDERAL FLUMINENSE
ESCOLA DE ENGENHARIA
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA ELÉTRICA
E DE TELECOMUNICAÇÕES

RAFAEL SOUZA AZEVEDO

**Aplicação de Modelos Ocultos de Markov na Caracterização
Banda Larga do Canal Rádio Móvel em Diferentes Ambientes**

Dissertação de Mestrado apresentada ao Programa de Pós-Graduação em Engenharia Elétrica e de Telecomunicações da Universidade Federal Fluminense, como requisito parcial para obtenção do título de Mestre em Engenharia Elétrica e de Telecomunicações. Área de concentração: Sistemas de Telecomunicações.

Orientadora: Prof^a. Dra. LENI JOAQUIM DE MATOS
Coorientador: Prof^o Dr. EDSON LUIZ CATALDO FERREIRA

NITERÓI

2023

Ficha catalográfica automática - SDC/BEE
Gerada com informações fornecidas pelo autor

A994a Azevedo, Rafael Souza
Aplicação de modelos ocultos de Markov na caracterização
banda larga do canal rádio móvel em diferentes ambientes /
Rafael Souza Azevedo. - 2023.
133 f.: il.

Orientador: Leni Joaquim de Matos.
Coorientador: Edson Luiz Cataldo Ferreira.
Dissertação (mestrado)-Universidade Federal Fluminense,
Escola de Engenharia, Niterói, 2023.

1. Canal rádio móvel. 2. Modelos ocultos de Markov. 3.
Sondagem em faixa larga. 4. Dispersão temporal. 5. Produção
intelectual. I. Matos, Leni Joaquim de, orientadora. II.
Ferreira, Edson Luiz Cataldo, coorientador. III. Universidade
Federal Fluminense. Escola de Engenharia.IV. Título.

CDD - XXX

RAFAEL SOUZA AZEVEDO

Aplicação de Modelos Ocultos de Markov na Caracterização
Banda Larga do Canal Rádio Móvel em Diferentes Ambientes

Dissertação de Mestrado apresentada ao Programa de Pós-Graduação em Engenharia Elétrica e de Telecomunicações da Universidade Federal Fluminense, como requisito parcial para obtenção do título de Mestre em Engenharia Elétrica e de Telecomunicações. Área de concentração: Sistemas de Telecomunicações.

BANCA EXAMINADORA

Prof.^a Leni Joaquim de Matos, D.Sc. – Orientador, UFF

Prof. Edson Luiz Cataldo Ferreira, D.Sc. – Coorientador, UFF

Prof. Pedro Vladimir Gonzalez Castellanos, D.Sc. – UFF

Prof. Carlos Vinicio Rodríguez Ron, D.Sc. – PUC-RJ

Niterói

Março de 2023

Dedico este trabalho a minha querida esposa, presente em todos os momentos.

Dedico, também, aos meus pais. Pilares da minha formação como ser humano.

Agradecimentos

À minha esposa, Priscilla, por todo amor, companheirismo, compreensão, incentivo, paciência e carinho.

Aos meus pais e aos meus irmãos pelo apoio e incentivo que serviram de alicerce para as minhas realizações.

Aos meus orientadores, professores Leni Joaquim de Matos e Edson Luiz Cataldo Ferreira, por guiarem meus passos ao longo dessa jornada.

À coordenação do PPGEET, por toda a ajuda dispensada.

Resumo

Atualmente, o ecossistema criado em torno dos sistemas de comunicações móveis produz enormes contribuições tecnológicas que geram inúmeros empregos diretos e indiretos e trazem incontáveis melhorias em produtividade e eficiência. Os impactos sociais e econômicos são enormes, uma vez que a conectividade está sempre presente, permitindo acesso à informação em tempo real através de ferramentas e plataformas digitais que permeiam o cotidiano dos cidadãos. Toda essa evolução tem permitido inúmeras e inesperadas inovações. O apetite por toda essa oferta de bens e serviços é voraz e tende a crescer em comum acordo com o próprio desenvolvimento e a exploração de novas tecnologias.

O aperfeiçoamento desses sistemas não engloba somente melhorias em dispositivos móveis, exploração de novos modelos de negócios, surgimento de novas tecnologias, melhorias em segurança da informação, mas abrangem também análises mais apuradas e estudos robustos dos ambientes de propagação do sinal rádio. Esses meios de propagação são grandes limitadores de desempenho dos sistemas de comunicações móveis e a adequada capacidade de predição da propagação do sinal rádio nesses meios é crucial.

Em vista disso, essa dissertação objetiva modelar o canal rádio móvel banda larga em diferentes ambientes, empregando modelos ocultos de Markov aos principais parâmetros de dispersão temporal do canal. Ao viabilizar tal modelagem, o uso da técnica empregada mostra-se como uma nova e poderosa opção para a caracterização estatística do canal rádio móvel.

Palavras-chaves: Canal rádio móvel, canal banda larga, modelos ocultos de Markov, topologia de Bakis, varredura em frequência, parâmetros de dispersão temporal do canal.

Abstract

Currently, the ecosystem created around mobile communication systems produces huge technological contributions that generate countless direct and indirect jobs and bring numerous improvements in productivity and efficiency. The social and economic impacts are enormous, since connectivity is always present, allowing access to information in real time through digital tools and platforms that permeate the daily lives of the citizens. All this evolution has allowed unexpected and incalculable number of innovations. The appetite for all this supply of goods and services is voracious and tends to grow in common with the development and exploitation of new technologies.

The improvement of these systems not only encompasses improvements in mobile devices, exploration of new business models, emergence of new technologies, improvements in information security, but also encompasses more accurate analyzes and robust research of radio signal propagation environments. These radio propagation environments place huge and fundamental limitations on the performance of radio communications systems and the adequate ability to predict the propagation of the radio signal in these environments is crucial.

Therefore, this thesis aims to model the broadband mobile radio channel in different environments, using hidden Markov models to the main temporal dispersion parameters of the channel. When proving the feasibility of such modelling, the employed technique is shown to be a new and powerful option for the statistical characterization of the radio channel.

Keywords: Mobile radio channel, broadband channel, hidden Markov models, Bakis topology, swept frequency (chirp), temporal dispersion parameters of the channel.

Lista de Figuras

FIGURA 1- MODELO DO CANAL DE COMUNICAÇÃO.	8
FIGURA 2- REPRESENTAÇÃO DE UMA VARIÁVEL ALEATÓRIA.	10
FIGURA 3- REPRESENTAÇÃO DOS MECANISMOS DE PROPAGAÇÃO.	12
FIGURA 4- VARIABILIDADE DA POTÊNCIA DO SINAL AO LONGO DO PERCURSO DE MEDIÇÃO.	13
FIGURA 5- FERRAMENTA PARA PROJETOS DE SISTEMAS DE COMUNICAÇÃO RÁDIO MÓVEL [23].	14
FIGURA 6- (A) DESVANECIMENTO EM GRANDE ESCALA AO LONGO DO PERCURSO. (B) MESMO DESVANECIMENTO EM ESCALA LOGARÍTMICA. SEPARAÇÃO EM PERDA DE PERCURSO E SOMBREAMENTO. (C) REMOÇÃO DA PERDA DE PERCURSO E VISUALIZAÇÃO DO EFEITO DE SOMBREAMENTO (D) SOMBREAMENTO APRESENTA NATUREZA LOG-NORMAL.	15
FIGURA 7- DESVANECIMENTO EM PEQUENA ESCALA EM AMBIENTE <i>OUTDOOR</i>	17
FIGURA 8- POTÊNCIA DO SINAL MEDIDO EM UM PONTO EM AMBIENTE INTERNO (<i>INDOOR</i>).	17
FIGURA 9- REPRESENTAÇÃO DO HMM.	27
FIGURA 10- APARELHO FONADOR HUMANO.	32
FIGURA 11- ESCALA MEL.	33
FIGURA 12- TOPOLOGIA ERGÓDICA.	35
FIGURA 13- TOPOLOGIA ESQUERDA-DIREITA.	36
FIGURA 14- TOPOLOGIA DE BAKIS.	37
FIGURA 15- TOPOLOGIA LINEAR.	37
FIGURA 16- (A) PLANTA DO AMBIENTE A5. (B) VISTA DO AMBIENTE A5.	41
FIGURA 17- (A) PLANTA DO AMBIENTE A6. (B) VISTA DO AMBIENTE A6.	42
FIGURA 18- (A) PLANTA DO AMBIENTE A7. (B) VISTA DO AMBIENTE A7.	42
FIGURA 19- (A) PLANTA DO AMBIENTE A8. (B) VISTA DO AMBIENTE A8.	43
FIGURA 20- (A) PLANTA DO AMBIENTE B3. (B) VISTA DO AMBIENTE B3.	44
FIGURA 21- (A) PLANTA DO AMBIENTE B4. (B) VISTA DO AMBIENTE B4.	44
FIGURA 22- (A) PLANTA DO AMBIENTE C1. (B) VISTA DO AMBIENTE C1.	45
FIGURA 23- (A) PLANTA DO AMBIENTE C4. (B) VISTA DO AMBIENTE C4.	45
FIGURA 24- AMBIENTE DE MEDIÇÃO A5 [38].	51
FIGURA 25- PLANTA DETALHADA DO AMBIENTE DE MEDIÇÃO A5.	53
FIGURA 26- PRIMEIRO PONTO DE RX1 PARA O AMBIENTE DE MEDIÇÃO A5.	54
FIGURA 27- FUNÇÃO JANELA A5 BLACKMAN-HARRIS APLICADA NO MATLAB.	54
FIGURA 28- APLICAÇÃO DA TÉCNICA CFAR PARA LIMPEZA DE RUÍDO.	55
FIGURA 29- REPRESENTAÇÃO DA MATRIZ DE CARACTERÍSTICAS POR LOCALIDADE DE RX1.	56
FIGURA 30- ADOÇÃO DA TOPOLOGIA DE BAKIS PARA TRANSIÇÃO ENTRE ESTADOS.	57
FIGURA 31- REPRESENTAÇÃO DA MATRIZ DE CARACTERÍSTICAS DE TRANSIÇÃO.	58
FIGURA 32- ADOÇÃO DO HMM E DA TOPOLOGIA DE BAKIS PARA CARACTERIZAÇÃO DO CRM.	58
FIGURA 33- REPRESENTAÇÃO DA MATRIZ BASE DE CARACTERÍSTICAS.	59
FIGURA 34- REPRESENTAÇÃO DOS PASSOS PARA APLICAÇÃO DO HMM AO CRM.	59
FIGURA 35- ESTIMAÇÃO DE F.D.P. PARA A TRANSIÇÃO 3-4 DO AMBIENTE A5.	63
FIGURA 36- ESTIMAÇÃO DE F.D.P. PARA 9 TRANSIÇÕES DO AMBIENTE A5.	65
FIGURA 37- ESTIMAÇÃO DE F.D.P. PARA 9 TRANSIÇÕES DO AMBIENTE A6.	67
FIGURA 38- ESTIMAÇÃO DE F.D.P. PARA 3 TRANSIÇÕES DO AMBIENTE A7.	69
FIGURA 39- ESTIMAÇÃO DE F.D.P. PARA 3 TRANSIÇÕES DO AMBIENTE A8.	70
FIGURA 40- ESTIMAÇÃO DE F.D.P. PARA 3 TRANSIÇÕES DO AMBIENTE C4.	72
FIGURA 41- ESTIMAÇÃO DE F.D.P. PARA 4 TRANSIÇÕES DO AMBIENTE B4.	74
FIGURA 42- ESTIMAÇÃO DE F.D.P. PARA 4 TRANSIÇÕES DO AMBIENTE C1.	75
FIGURA 43- ESTIMAÇÃO DE F.D.P. PARA 5 TRANSIÇÕES DO AMBIENTE B3.	77
FIGURA 44- ESTIMAÇÃO DE F.D.P. PARA A5 COMO TREINAMENTO E A6 PARA TESTE E VALIDAÇÃO.	79
FIGURA 45- ESTIMAÇÃO DE F.D.P. PARA A5 COMO TREINAMENTO E C1 PARA TESTE E VALIDAÇÃO.	82

Lista de Tabelas

TABELA 1 - PARÂMETROS DO SISTEMA DE SONDAÇÃO.....	49
TABELA 2 - ESPECIFICAÇÃO DOS LIMIARES ESCOLHIDOS [21].	61
TABELA 3 - MATRIZ DE CONFUSÃO PARA CADA TRANSIÇÃO DE ESTADOS DE A5.	63
TABELA 4 - MATRIZ DE CONFUSÃO PARA O CONJUNTO DE TRANSIÇÕES DE A5.....	66
TABELA 5 - MATRIZ DE CONFUSÃO PARA O CONJUNTO DE TRANSIÇÕES DE A6.....	67
TABELA 6 - MATRIZ DE CONFUSÃO PARA O CONJUNTO DE TRANSIÇÕES DE A7.....	69
TABELA 7 - MATRIZES DE CONFUSÃO PARA O CONJUNTO DE TRANSIÇÕES DE A8.....	71
TABELA 8 - MATRIZES DE CONFUSÃO PARA O CONJUNTO DE TRANSIÇÕES DE C4.....	72
TABELA 9 - MATRIZ DE CONFUSÃO PARA O CONJUNTO DE TRANSIÇÕES DE B4.....	74
TABELA 10 - MATRIZ DE CONFUSÃO PARA O CONJUNTO DE TRANSIÇÕES DE C1.....	76
TABELA 11 - MATRIZ DE CONFUSÃO PARA O CONJUNTO DE TRANSIÇÕES DE B3.....	78
TABELA 12 - MATRIZ DE CONFUSÃO PARA A SIMULAÇÃO COM OS AMBIENTES A5 E A6.	80
TABELA 13 - MATRIZ DE CONFUSÃO PARA A SIMULAÇÃO COM AMBIENTES A5 E C1.....	82
TABELA 14 – AGRUPAMENTO DE CARACTERÍSTICAS E RESULTADOS DAS SIMULAÇÕES.	86

Lista de Abreviaturas e Siglas

3GPP	<i>3rd Generation Partnership Project</i>
5G NSA	<i>5G Non-Standalone</i>
5G SA	<i>5G Standalone</i>
AMPS	<i>Advanced Mobile Phone Service</i>
AR	<i>Augmented Reality</i>
CA	<i>Carrier Aggregation</i>
CFAR	<i>Constant False Alarm Rate</i>
CRM	Canal rádio móvel
CS	<i>Circuit Switch</i>
EDGE	<i>Enhanced Data rates for GSM Evolution (EDGE)</i>
EM	<i>Expectation-Maximization</i>
eMBB	<i>enhanced Mobile Broadband</i>
ETSI	<i>European Telecommunications Standards Institute</i>
FDD	<i>Frequency Division Duplex</i>
f.d.p.	Função de densidade de probabilidade
FFT	<i>Fast Fourier Transform</i>
FR	<i>Frequency Range</i>
GPRS	<i>General Packet Radio Service</i>
GSM	<i>Global System for Mobile Communications</i>
HMM	<i>Hidden Markov Model</i>
HPIB	<i>Hewlett Packard Interface Bus</i>
HSPA	<i>High Speed Packet Access</i>
HSPA+	<i>Evolved High Speed Packet Access</i>

ITU	<i>International Telecommunication Union</i>
IDFT	<i>Inverse Discrete Fourier Transform</i>
IoT	<i>Internet of Things</i>
IP	<i>Internet Protocol</i>
LoS	<i>Line of sight</i>
LTE	<i>Long-Term Evolution</i>
MFCCs	<i>Mel-Frequency Cepstral Coefficients</i>
MIMO	<i>Multiple Input Multiple Output</i>
mMTC	<i>massive Machine-Type Communication</i>
mmWaves	<i>millimeter Waves</i>
NLoS	<i>Non-Line of Sight</i>
NMT	<i>Nordic Mobile Telephone</i>
PPGEET	Programa de Pós Graduação em Engenharia Elétrica de Telecomunicações
PS	<i>Packet Switch</i>
QWSSUS	<i>Quase-Wide-Sense Stationary Uncorrelated Scattering</i>
SDOs	<i>Standard Development Organizations</i>
UMTS	<i>Universal Mobile Telecommunications System</i>
URLLC	<i>Ultra-Reliable Low Latency Communications</i>
US	<i>Uncorrelated Scattering</i>
TDD	<i>Time Division Duplex</i>
V2X	<i>Vehicle-to-everything</i>
v.a.	Variável aleatória
VR	<i>Virtual Reality</i>
WSS	<i>Wide-Sense Stationary</i>
WSSUS	<i>Wide-Sense Stationary Uncorrelated Scattering</i>

Sumário

Capítulo 1 - Introdução	1
1.1 Contexto histórico	1
1.2 Motivação.....	4
1.3 Objetivos	6
1.4 Estrutura do Trabalho	6
Capítulo 2 - Conceituação Teórica.....	8
2.1 Representação de fenômenos por meio de modelos	9
2.2 Canal rádio móvel (CRM).....	10
2.2.1 Desvanecimento (<i>fading</i>).....	13
2.2.1.1 Desvanecimento em grande escala.....	14
2.2.1.1 Desvanecimento em pequena escala	16
2.3 Caracterização do canal rádio móvel	18
2.3.1 Caracterização faixa estreita.....	18
2.3.2 Caracterização faixa larga.....	18
2.3.2.1 Caracterização do canal rádio real.....	19
2.3.2.2 Técnicas de sondagem do canal banda larga.....	20
2.4 Parâmetros de dispersão do canal rádio móvel.....	20
2.4.1 Parâmetros de dispersão no tempo.....	21
2.4.1.1 Retardo médio (<i>average delay</i>)	21
2.4.1.2 Espalhamento de retardo (<i>delay spread</i>)	21
2.4.1.3 Banda de coerência (<i>coherence bandwidth</i>).....	22
2.4.2 Parâmetros de dispersão na frequência	22
2.4.2.1 Deslocamento Doppler	22
2.4.2.2 Espalhamento Doppler (<i>Doppler spread</i>).....	22
2.4.2.3 Tempo de coerência (<i>coherence time</i>).....	23
2.5 Modelos de Markov e cadeias de Markov	23
2.6 Modelos ocultos de Markov (HMM).....	25
2.7 Problemas controle do HMM	29
2.7.1 Problema de avaliação (<i>evaluation problem</i>).....	30
2.7.2 Problema de decodificação (<i>decoding problem</i>).....	30
2.7.3 Problema de aprendizagem e treinamento (<i>learning problem</i>).....	30
2.8 Comparação da modelagem dos parâmetros do canal com a identificação de locutor através do uso de HMM.....	31

2.9	Topologias para transição de estados.....	34
2.9.1	Topologia ergódica (<i>ergodic</i>)	35
2.9.2	Topologia esquerda-direita (<i>left-to-right</i>).....	36
2.9.3	Topologia de Bakis	36
2.9.4	Topologia linear	37
Capítulo 3 - Sondagem do canal rádio e cenários de medições		39
3.1	Adoção da técnica de varredura em frequência.....	39
3.2	Campanha de medições	40
3.2.1	Ambientes da campanha de medições	40
3.2.1.1	Ambiente de medição A5 (<i>indoor-outdoor</i>)	41
3.2.1.2	Ambiente de medição A6 (<i>outdoor-indoor</i>)	42
3.2.1.3	Ambiente de medição A7 (<i>indoor-outdoor</i>)	42
3.2.1.4	Ambiente de medição A8 (<i>outdoor-indoor</i>)	43
3.2.1.5	Ambiente de medição B3 (<i>outdoor</i>).....	43
3.2.1.6	Ambiente de medição B4 (<i>outdoor</i>).....	44
3.2.1.7	Ambiente de medição C1 (<i>outdoor</i>).....	45
3.2.1.8	Ambiente de medição C4 (<i>outdoor</i>).....	45
3.2.2	Equipamentos da campanha de medições.....	46
3.2.3	Preparação (<i>setup</i>) dos equipamentos e ambientes	47
3.2.4	Coleta oficial de dados	50
Capítulo 4 - Metodologia proposta		52
4.1	Adoção do canal como QWSSUS.....	52
4.2	Aquisição e pré-processamento dos dados.....	53
4.3	Adoção da topologia de Bakis	56
4.4	Caracterização do CRM com uso do HMM.....	57
Capítulo 5 - Simulações e resultados		62
5.1	Simulação 1: transições aleatórias entre os estados do mesmo ambiente de medição A5 (<i>indoor-outdoor</i>)	62
5.2	Simulação 2: transições determinadas entre os estados do mesmo ambiente de medição A5 (<i>indoor-outdoor</i>)	64
5.3	Simulação 3: transições determinadas entre os estados do mesmo ambiente de medição A6 (<i>outdoor-indoor</i>)	66
5.4	Simulação 4: transições determinadas entre os estados do mesmo ambiente de medição A7 (<i>indoor-outdoor</i>)	68
5.5	Simulação 5: transições determinadas entre os estados do mesmo ambiente de medição A8 (<i>outdoor-indoor</i>)	70
5.6	Simulação 6: transições determinadas entre os estados do mesmo ambiente de medição C4 (<i>outdoor</i>)	71

5.7	Simulação 7: transições determinadas entre os estados do mesmo ambiente de medição B4 (<i>outdoor</i>).....	73
5.8	Simulação 8: transições determinadas entre os estados do mesmo ambiente de medição C1 (<i>outdoor</i>).....	75
5.9	Simulação 9: transições determinadas entre os estados do mesmo ambiente de medição B3 (<i>outdoor</i>).....	76
5.10	Simulação 10: transições determinadas entre os estados com grupo de treinamento e grupo de teste de distintos ambientes de medição A5 (<i>indoor-outdoor</i>) e A6 (<i>outdoor-indoor</i>).....	78
5.11	Simulação 11: transições determinadas entre os estados com grupo de treinamento e grupo de teste de distintos ambientes de medição A5 (<i>outdoor-indoor</i>) e C1 (<i>outdoor</i>).....	80
5.12	Resultados	83
5.12.1	Primeira etapa: medições realizadas	84
5.12.2	Segunda etapa: pré-processamento dos dados	84
5.12.3	Terceira etapa: caracterização do CRM com uso do HMM.....	85
Capítulo 6 - Conclusões e trabalhos futuros		87
6.1	Conclusões	87
6.2	Sugestões para trabalhos futuros	89
Bibliografia		90
Anexo A – Programa em Matlab: pré-processamento dos dados		93
Anexo B – Programa em Matlab: criação da matriz base de características		101
Anexo C – Programa em Matlab: Processamento dos dados com HMM <i>Toolbox</i>		110

Capítulo 1 - Introdução

1.1 Contexto histórico

No início do século XX, Nikola Tesla descreveu princípios de aplicações contemporâneas de sistemas de comunicações via rádio. Ele contou que seria possível ditar instruções entre pontos distantes, usar um pequeno e barato instrumento que permitira ao portador receber e ouvir, em qualquer lugar, no mar ou na terra, músicas, discursos, sermões e que imagens poderiam ser transferidas entre quaisquer locais [1]. As previsões de Tesla não foram ao acaso, visto que ele teve ativa participação na vanguarda das pesquisas que já vinham se desenvolvendo desde o final do século anterior.

No final do século XIX, seguindo trabalhos pioneiros como os trabalhos de James C. Maxwell e Heinrich R. Hertz sobre ondas eletromagnéticas, Oliver Lodge demonstrou a viabilidade da comunicação radio em curta distância e coube ao empreendedor Guglielmo Marconi demonstrar a viabilidade da comunicação rádio em longas distâncias. Entre vários experimentos bem sucedidos, Marconi estabeleceu com sucesso a comunicação rádio entre uma estação em terra e um barco rebocador, localizado alguns quilômetros de distância do litoral. Embora o objetivo não fosse o estabelecimento de uma comunicação rádio móvel, tal experimento certamente pode ser caracterizado como tal.

Desde então, sistemas de comunicações rádio móveis têm-se desenvolvido continuamente ao redor do mundo. Como a propagação da onda eletromagnética não se preocupa com fronteiras geopolíticas, o estabelecimento de acordos internacionais foi e continua sendo importante para a propícia exploração do espectro de frequência e, nesse contexto, para facilitar a conectividade internacional em redes de comunicações, destaca-se a *International Telecommunication Union* (ITU) como a principal agência para firmar equidade e harmonia global.

Os princípios das comunicações rádio móveis terrestres remontam à década de 20, quando a polícia de Detroit iniciou o uso de sistemas unidirecionais de comunicação rádio móvel para transmitir informações para as suas viaturas. As décadas seguintes apresentaram novos desenvolvimentos e novas implementações, até que na década de 80 os primeiros sistemas de telefonia celular foram inaugurados em diferentes cenários globais: o *Nordic Mobile Telephone* (NMT) em 1981 na Escandinávia, o *Advanced Mobile*

Phone Service (AMPS) em 1983 nos Estados Unidos e o *Global System for Mobile Communications* (GSM) em 1988 na Europa [2].

O GSM nasceu em um período de grande cooperação europeia, o seu desenvolvimento foi conduzido pelo *European Telecommunications Standards Institute* (ETSI) e tinha como propósito a substituição dos sistemas analógicos de comunicações móveis europeus existentes, visto que eles eram incompatíveis entre si. Ele se caracteriza por já ser um sistema digital, qualificando-o como uma segunda geração em relação aos antecessores sistemas de primeira geração que eram analógicos.

Os lançamentos e as evoluções contínuas dos sistemas de comunicações móveis celulares ao longo das últimas décadas seguem em acordo com as necessidades de consumo mundiais e foram impulsionadas, na década de 90, pela popularização da internet. Sistemas móveis celulares, que nasceram com o intuito de proporcionar comunicação por voz, estavam diante de uma constante e crescente demanda por disponibilidade de acesso a informação. As pessoas naturalizaram o fato de possuírem dispositivos dedicados à comunicação e foram além, elas desejavam mais do que falar, cada vez mais ansiavam pelo consumo de informação instantânea. Essa acelerada demanda por informação fez crescer vertiginosamente o número de dispositivos móveis celulares usados ao redor do mundo (o número de assinantes móveis já ultrapassou a casa de 5,4 bilhões em 2022) [3] e, com esse crescimento, somaram-se novos e variados serviços e maiores demandas inquietas pelo consumo de informação.

O GSM, que nasceu para dar suporte à voz entre dois usuários via comutação de circuito (*Circuit Switch*, CS), incorporou novas tecnologias como o *General Packet Radio Service* (GPRS) e posteriormente o *Enhanced Data rates for GSM Evolution* (EDGE) para prover transmissão de dados via comutação de pacotes (*Packet Switch*, PS).

A evolução natural do GSM, dado o aumento constante da demanda por conectividade, ocorreu no início do século XXI. Uma vez que o GSM teve grande alcance, os empenhos para o desenvolvimento de um novo sistema global de comunicações móveis migraram do ETSI para um projeto global que reuniu esforços de diferentes organizações de desenvolvimento de padrões de telecomunicações, chamadas de *Standard Development Organizations* (SDOs). Esse projeto foi denominado *3rd Generation Partnership Project* (3GPP) e ele definiu um sistema de comunicação móvel global de terceira geração, denominado *Universal Mobile Telecommunications System* (UMTS). Essa terceira geração já garantia a capacidade de conectividade via CS e PS, provendo uma melhor

conectividade com a internet. Posteriormente ainda foram incorporadas novas tecnologias como o *High Speed Packet Access* (HSPA) e o *Evolved High Speed Packet Access* (HSPA+), que melhoraram ainda mais a conectividade.

A geração seguinte dos sistemas de comunicações móveis celulares foi lançada no final da década de 2000 e tomou como base uma filosofia que não se baseou na rede legada até então desenvolvida. Dessa forma, a definição de requisitos e metas de configuração da nova rede não se restringiu aos trabalhos anteriormente adotados. Isso possibilitou um projeto totalmente novo, onde a rede foi completamente otimizada para suportar PS, não havendo a exigência de amparar CS, requisito esse que a terceira geração possuía. Embora não houvesse exigência, houve o amparo por parte das operadoras de telefonia móvel com o intuito de continuar usando a rede legada. Dessa forma, o UMTS evoluiu para a quarta geração dos sistemas de comunicações móveis celulares, denominada de *Long-Term Evolution* (LTE), ao se enquadrar em requisitos como: rede com arquitetura PS completamente baseada em IP (*all-IP*, all *Internet Protocol*), altas taxas de dados nas bordas das células, taxas de pico de dados aumentadas significativamente, redução de latência para melhorar a usabilidade, flexibilidade de uso do espectro e a grande semelhança entre operações em *Frequency Division Duplex* (FDD) e *Time Division Duplex* (TDD) [4].

Em vista da evolução contínua dos sistemas de comunicações móveis celulares, parâmetros como a taxa de transmissão de dados e a latência são aprimorados a cada geração. Esses sistemas de alto desempenho evoluíram e continuam evoluindo para suportarem uma vasta variedade de dispositivos e serviços. Os atuais sistemas de telefonia móvel celular, cada vez mais, usam combinações de frequências e novas tecnologias. A atual quinta geração dos sistemas de comunicações móveis celulares (formalmente denominada IMT-2020), para lidar com diferentes cenários de implementação e ofertar uma vasta gama de serviços, utiliza arquitetura PS *all-IP*, pode utilizar faixas de frequências mais baixas (denominadas *Frequency Range 1* – FR1 ou Sub 6 GHz, com larguras de banda de até 100 MHz) ou empregar frequências mais altas (denominadas *Frequency Range 2* – FR2 ou *millimeter Waves* – *mmWave* ou ondas milimétricas, com larguras de banda de até 400 MHz) [5] [6] [7]. Soma-se a isso o fato dessa nova tecnologia ainda usar agregações espectrais (*Carrier Aggregation*, CA) e múltiplas antenas (*Multiple Input Multiple Output*, MIMO), combinando dinamicamente diferentes faixas de

frequências para aumentar a banda de transmissão e ganhar eficiência, desempenho e suprir as necessidades de diferentes segmentos de mercado.

A quinta geração expande o conceito de sistemas de comunicações móveis celulares para além dos costumes tradicionais. No final da década passada, o 3GPP expôs duas opções primárias de arquitetura que representam evoluções da rede LTE: 5G *Non-Standalone* (5G NSA, 3GPP Release 15) e 5G *Standalone* (5G SA, 3GPP Release 16). O 5G NSA permite uma implantação rápida e menos custosa, visto que aproveita a infraestrutura da rede LTE existente. Ao passo que o 5G SA compreende uma rede completamente nova, não necessitando da rede legada para prover serviços e fornecer aprimoramentos. A verdade é que a nova quinta geração vai além do fornecimento de taxas de transmissão de dados maiores e menor latência e suas oportunidades de negócio podem ser baseadas em três pilares: *enhanced Mobile Broadband* (eMBB) com foco em altas taxas de transmissão de dados para suportar serviços como realidade virtual (*virtual reality*, VR) e realidade aumentada (*augmented reality*, AR); *Ultra-Reliable Low Latency Communications* (URLLC) com foco em latência extremamente baixa para suportar serviços como direção autônoma, *vehicle-to-everything* (V2X) e cirurgia remota; *massive Machine-Type Communication* (mMTC) com foco em alta densidade de conectividade para suportar serviços como conceitos de cidades inteligentes e IoT massivo (*massive Internet of Things*) [6]. Com base nos pilares acima, as operadoras de telefonia móvel são capazes de ofertar serviços julgados por elas como sendo necessários para atender as novas necessidades do mercado.

É admirável olhar para o passado e se deparar com os trabalhos de James C. Maxwell e Heinrich R. Hertz, com as previsões acertadas de Nikola Tesla e com os resultados alcançados por Guglielmo Marconi. Bem como é surpreendente observar a contínua evolução dos sistemas de comunicações móveis, verificar o quanto já evoluiu e o quanto essa própria evolução nutre novas pesquisas, promissores resultados e a contínua e crescente demanda global por produtos e serviços.

1.2 Motivação

Ao passo do atual estágio de desenvolvimento dos sistemas de comunicações móveis, o aperfeiçoamento de tais sistemas engloba não apenas melhorias em dispositivos

móveis e uso de novas tecnologias, mas abrange também análises mais apuradas dos meios de propagação e do adequado dimensionamento dos canais rádio móveis.

Os meios de propagação apresentam-se como grandes limitadores do desempenho de sistemas de comunicações rádio móveis. Até então, visto que é mais proveitoso trabalhar com modelos que descrevem a propagação do sinal do que trabalhar com dados e medições da propagação em si, posto também a complexidade dos fenômenos naturais envolvidos e os custos inerentes às medições, desenvolveram-se métodos para mitigar as limitações impostas pelos meios de propagação e para prover adequadas caracterizações dos canais rádio móveis.

Tais caracterizações já são amplamente estudadas e difundidas. Cabe aqui, citar alguns consagrados modelos físicos de modelagem faixa larga do canal rádio móvel (CRM) como o modelo de Saleh-Valenzuela [8] [9], que caracteriza o comportamento dispersivo do canal rádio *indoor* no domínio do tempo através da geração sucessiva de respostas instantâneas; modelo auto-regressivo (*autoregressive model*) [10] [11], que calcula os parâmetros do modelo a partir de medições realizadas de resposta do canal na frequência e emprega o princípio da máxima entropia; e modelo de geração e desaparecimento de raios (*Generation and Disappearance of Paths*) [12] [13], que descreve estatisticamente os parâmetros (amplitude, fase e retardo) associados a cada multipercurso gerado pelo canal no domínio do tempo.

Em vista do exposto acima, a motivação para a presente dissertação foi prosseguir com o trabalho inicialmente desenvolvido por Ribeiro [14]. Esse trabalho iniciou um estudo da aplicação de modelos ocultos de Markov (*Hidden Markov Model*, HMM) na caracterização faixa larga do canal rádio, trabalhando com os parâmetros temporais de retardo médio e espalhamento de retardo, obtidos do processamento de medidas realizadas em ambiente *indoor*, chegando a bons resultados para os valores simulados. Dessa maneira, o emprego do HMM, na modelagem estatística dos parâmetros do canal, traz uma nova forma de modelagem, que tem satisfatória aplicação em sinais que variam com o tempo, pois endossa, em sua natureza probabilística e adaptativa, a maior vantagem para caracterizar sinais corrompidos por ruído, tais como a fala, a escrita e o próprio sinal rádio. A ideia para o emprego de tal modelagem foi inspirada no uso do método HMM para identificação de um particular indivíduo, denominado de locutor, no qual se observam os estados como fonemas de uma palavra [15] [16] [17] [18] [19] [20].

1.3 Objetivos

Ainda que certos fenômenos não sejam completamente perceptíveis, provavelmente eles geram observáveis que ajudam na modelagem e, conseqüentemente, na compreensão do próprio fenômeno em si.

Ribeiro [21] definiu pares de regiões de medições e a transição entre cada par de regiões de medições foi analisada individualmente e também comparada com as demais transições. Dessa forma, foi demonstrado que a técnica HMM ratifica as estatísticas de retardo médio e espalhamento de retardo do sinal para o CRM analisado, em ambiente *indoor*.

Nessa dissertação, investiga-se o CRM com o uso do método estatístico nomeado HMM. Objetiva-se expandir as primeiras simulações desenvolvidas por Ribeiro [21] e para tal, buscou-se explorar ambos os assuntos (HMM e CRM) de forma a listar as ideias envolvidas, conciliar similaridades, ajustar as medições obtidas e evidenciar conceitos, fazendo uso de HMM para caracterização e simulação estatística do CRM. As simulações não dizem respeito mais a uma única transição entre pares de regiões de medições, mas correspondem a um conjunto maior de transições entre regiões ou ao próprio ambiente de medição como um todo. Dessa forma, busca-se alcançar novos resultados para legitimar o uso do método HMM como uma nova opção para a modelagem do CRM.

1.4 Estrutura do Trabalho

Esta dissertação é composta por seis capítulos e está organizada seguinte forma:

- O capítulo 1 é um capítulo introdutório que busca contextualizar para o leitor o cenário no qual a presente dissertação está inserida. Ele mostra ainda, sucintamente, o objetivo e a estruturação da dissertação.
- O capítulo 2 descreve a conceituação teórica das ideias desenvolvidas posteriormente. Ao abranger a teoria sobre distintos tópicos como CRM e o método estatístico HMM, esse capítulo busca introduzir a teoria de ambos os temas e dos demais assuntos pertinentes. Além disso, ele objetiva estabelecer uma ligação entre os temas, de modo a facilitar a compreensão de tudo que foi desenvolvido.

- O capítulo 3 descreve os cenários de medições, bem como a técnica de sondagem adotada. Sobre as medições, buscou-se a divisão em tópicos e detalhamentos que retratassem, de forma clara, as passagens necessárias para a adequada aquisição das medições.
- O capítulo 4 expõe a metodologia proposta para a aplicação do método estatístico HMM na caracterização banda larga do CRM.
- O capítulo 5 trata das simulações e provê os respectivos resultados.
- O capítulo 6 discorre sobre as conclusões e os pertinentes direcionamentos para trabalhos futuros.

Capítulo 2 - Conceituação Teórica

Shannon [22] descreveu que o problema fundamental da comunicação consiste em reproduzir (totalmente ou parcialmente) no destino, uma mensagem selecionada na origem. Em vista disso, um canal de comunicação pode ser descrito como o caminho de conexão entre esses dois pontos [23]. Em outras palavras, o canal de comunicação representa o meio escolhido e utilizado para levar a mensagem, eleita na origem, até o destino. Esse canal pode ser representado por fios metálicos, fibras ópticas, ondas eletromagnéticas, pelo próprio ar como um meio mecânico (canal acústico) para propagação das ondas sonoras e da fala humana, etc. Ele conecta origem e destino e de forma geral, apesar de permitir a comunicação, acaba por interferir na mesma, adicionando interferências e ruídos.

O canal de comunicação, referenciado no presente estudo, diz respeito ao canal rádio. Pode-se dizer que o canal rádio engloba as antenas transmissora e receptora e o canal de propagação e, de certa forma, esses consistem em tudo que influencia na comunicação entre origem e destino, conforme a figura 1. Portanto, o canal de propagação, como esperado, corresponde ao meio físico entre as antenas transmissora e receptora e suporta as ondas eletromagnéticas que por ele propagam [23]. O canal rádio móvel (CRM), por sua vez, refere-se ao canal rádio que pode apresentar movimento relativo entre as antenas transmissora e receptora [24]. Veja que o movimento relativo apresenta-se de tal forma que o CRM é dito como um canal variante no tempo.

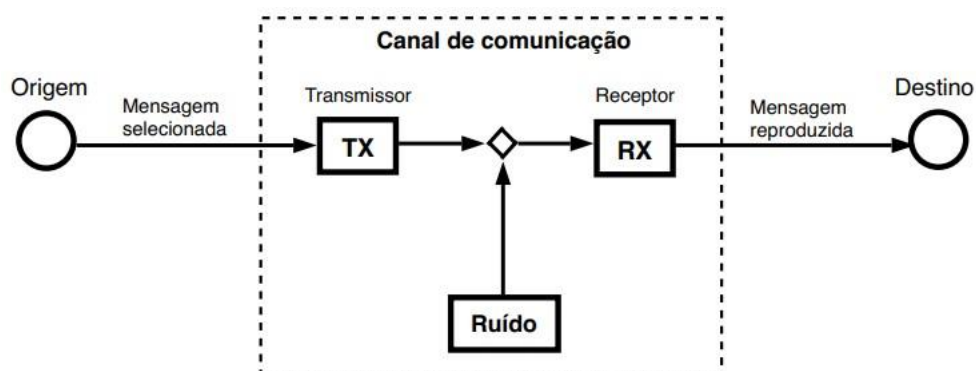


Figura 1- Modelo do canal de comunicação.

Um sistema de comunicação rádio móvel corresponde então ao aparato de equipamentos e dispositivos que permitem a criação e manutenção do CRM, de tal forma que os usuários desse sistema possam estabelecer uma comunicação sempre que

necessário. O desempenho desse sistema de comunicação sem fio depende, em grande parte, do seu respectivo ambiente de propagação. Dessa forma, é adequado afirmar que um particular canal de propagação influencia massivamente a comunicação entre as pertinentes antenas transmissora e receptora, de tal forma que esse canal de propagação, em grande medida, governa o desempenho do próprio sistema de comunicação rádio móvel [24] [25].

Diante disso, ao passo da ampliação contínua dos sistemas de comunicações móveis, com evoluções nas infraestruturas de rede e aperfeiçoamento dos dispositivos móveis, deve-se também desenvolver uma adequada análise do ambiente de propagação. Nesse contexto, o ambiente, referenciado como canal de propagação, apresenta-se como o canal de comunicação entre transmissor e receptor e o seu adequado dimensionamento possibilita uma comunicação mais eficiente.

2.1 Representação de fenômenos por meio de modelos

Em primeiro lugar, cabe uma descrição ampla e sucinta sobre princípios inerentes ao estudo em questão. De forma geral, fenômenos do mundo real podem ser demasiadamente complexos e, para que tais fenômenos sejam estudados, usam-se modelos que buscam a retratação dos processos envolvidos (do dinamismo envolvido). Essa representação, mesmo que simplista quando comparado à realidade, pode proporcionar uma adequada explicação do fenômeno estudado. Eventualmente, parâmetros, que foram omitidos no modelo adotado, podem aparecer como componentes desconhecidas, mas mesmo esses elementos inesperados e desconhecidos podem ser estruturados e agregados aos modelos adotados.

Os modelos podem ter diferentes finalidades, dependendo do objetivo em questão. Além de proporcionar uma descrição teórica que exponha o comportamento adequado do próprio fenômeno, como dito acima, eles podem ajudar na entrega de um resultado desejado ou proporcionar uma melhor compreensão da sua origem geradora [16].

Os modelos, também nomeados como processos, podem ser determinísticos, onde a mesma saída sempre será resultante da mesma entrada; aleatórios ou estocásticos, onde a saída representará uma probabilidade de ocorrência dada uma entrada conhecida; ou uma combinação de ambos os modelos anteriores.

Para a adoção de modelos com características aleatórias, torna-se necessária a introdução do conceito de variável aleatória (v.a.). Uma v.a., exemplificada aqui como Q ,

também chamada de variável estocástica, é uma função que assinala um número real q para cada possível saída s no espaço de amostras S de um experimento aleatório. Logo, como função, a v.a. Q corresponde a uma regra que atribui um valor numérico real para cada elemento do espaço amostral S . O espaço amostral S é o domínio da v.a. O conjunto S_Q é a imagem da v.a. e diz respeito a todos os valores obtido para Q no conjunto de estados S (espaço amostral S) [26]. A figura 2 mostra a representação da v.a. no conjunto real de valores.

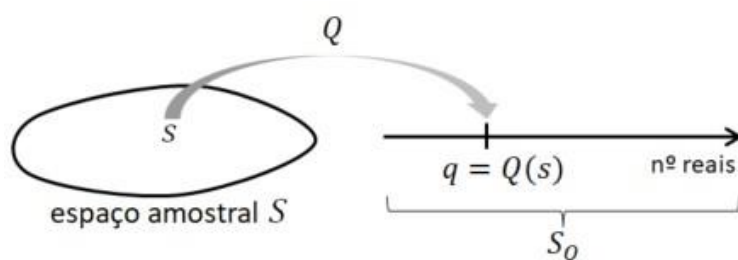


Figura 2- Representação de uma variável aleatória.

Dessa maneira, para diferentes momentos do experimento aleatório, tem-se a v.a. $Q(s_t) = q_t$, onde, no índice t , a v.a. Q está associando uma saída s_t do espaço amostral S a um valor real q_t . Assim, mesmo fenômenos complexos podem ser modelados.

Embora a inspeção da dinâmica que governa determinados fenômenos do mundo real não seja diretamente perceptível, provavelmente esses fenômenos geram observáveis, também denominados de sinais, os quais podem apresentar uma estrutura implícita que pode ser revelada por diferentes abordagens, como modelos Gaussianos, de Poisson, Markovianos, modelos ocultos de Markov, entre outros [16].

2.2 Canal rádio móvel (CRM)

Como dito anteriormente, o canal de comunicação adotado no presente estudo é o canal rádio e, dado a possível mobilidade relativa entre as antenas transmissora e receptora, esse canal é referenciado como móvel. Fenômenos, como o CRM, associam sinais de excitação (transmissão) e resposta (recepção) por meio de processos ditos estocásticos, representando sequências evolutivas de eventos com o tempo.

O ambiente entre as antenas transmissora e receptora, apresentado aqui como canal de propagação, faz parte do CRM e é descrito como o meio físico através do qual uma conexão é estabelecida para a propícia comunicação entre origem e destino. As ondas eletromagnéticas, abrangendo aqui as ondas de rádio, propagam-se através desse canal e são afetadas, principalmente, por fenômenos físicos e naturais como reflexão (*reflection*), refração (*refraction*), espalhamento (*scattering*) e difração (*diffraction*) [25]. Tais fenômenos são conhecidos como mecanismos de propagação. A reflexão ocorre quando a onda de rádio incide sobre um objeto com dimensões muito maiores se comparada ao comprimento da onda de rádio incidente (prédios, construções, etc.). De maneira geral, a reflexão pode ser acompanhada pela refração, que corresponde aos efeitos sofridos pela onda ao passar de um meio para outro diferente. As forças das ondas refletidas e refratadas dependem do material do respectivo objeto onde as ondas incidem. Já o espalhamento é relativo à incidência da onda de rádio sobre um meio com grande número de obstáculos em um pequeno espaço e onde as dimensões desses obstáculos (ditos espalhadores, *scatters*) são da ordem ou menores que o comprimento da onda de rádio incidente. Parsons [24] mostra que a difração diz respeito à curvatura aparente da onda de rádio ao encontrar quinas ou pequenas aberturas em obstáculos que se encontram entre transmissor e receptor, de tal forma que a onda parece contornar ou atravessar o respectivo obstáculo e continuar se propagando pelo meio.

Em um CRM, as ondas de rádio frequentemente não alcançam o destino diretamente. Muitas vezes existem muitos e diferentes obstáculos entre transmissor e receptor de tal forma que as ondas de rádio, em sua maioria, alcançam o destino via múltiplos percursos originados no canal de propagação. Isso caracteriza um canal sem visada direta (*non-line of sight*, NLoS). Em contrapartida, quando a maioria das ondas de rádio chega ao destino diretamente, quando existe majoritariamente uma linha de visada sem obstáculos entre origem e destino, isso caracteriza um canal com visada direta (*line-of-sight*, LoS) entre transmissor e receptor.

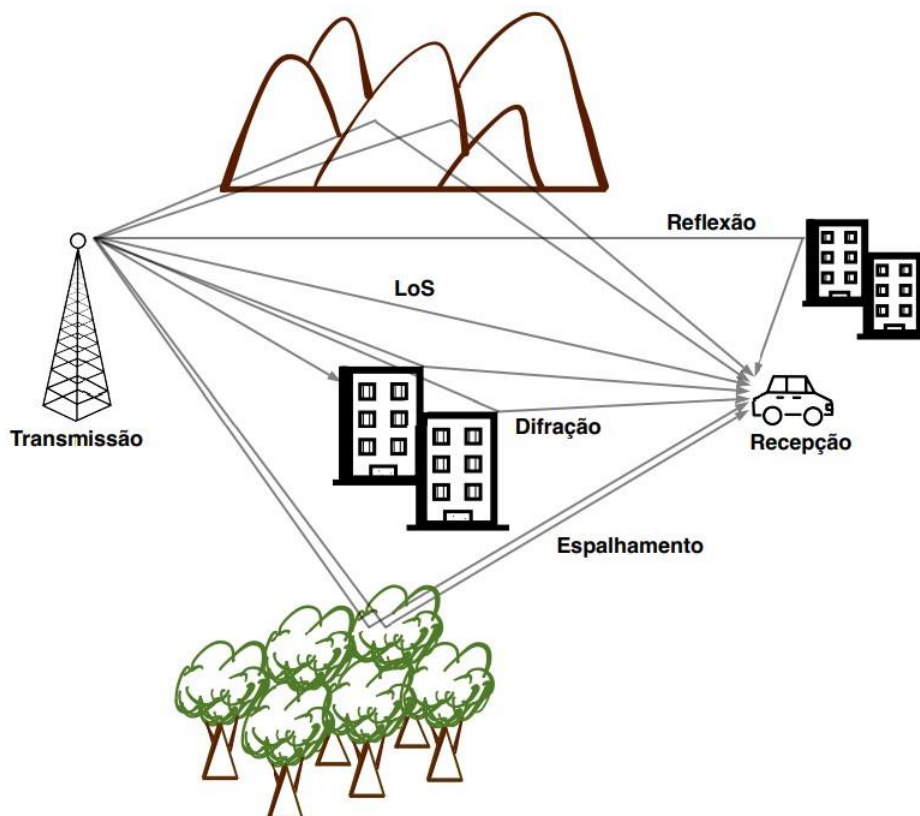


Figura 3- Representação dos mecanismos de propagação.

Em vista da ocorrência dos mecanismos de propagação acima citados, as ondas de rádio são frequentemente afetadas por efeitos reflexivos, refrativos, difrativos e espalhamentos que acabam por gerar múltiplas ondas que se propagam pelo ambiente até chegarem ao destino, conforme a figura 3. O receptor, nesse caso, encontra-se em um meio majoritariamente NLoS com vários obstáculos (montanhas, prédios, árvores, folhagens, carros, transeuntes, etc.), o que gera mais de um percurso para as ondas de rádio entre transmissão e recepção, de tal forma que o ambiente é dito como um ambiente multipercursos (*multipath environment*). O sinal recebido corresponde então ao somatório das múltiplas ondas, cada qual apresentando diferentes amplitudes, fases e atrasos, que se propagam pelos diferentes percursos (*multipath propagation*) e se superpõem no receptor.

Percebe-se, de fato, que um canal de propagação governa e limita o desempenho de um sistema de comunicação, de tal forma que diferentes canais (diferentes ambientes) apresentam distintos efeitos (construtivos ou destrutivos) dos mecanismos de propagação. Perante isso, análises das propagações tornam-se valiosas e objetivam estimar os comportamentos dos canais de propagação. O desenvolvimento de tais análises originaram as caracterizações dos canais rádio móvel (CRM), de maneira que esses canais são

atualmente representados por modelos estatísticos ou modelos que buscam descrever o comportamento do canal rádio em configurações e ambientes específicos como ambientes internos (*indoor*), externos (*outdoor*), ambientes de passagem interno-externo (*indoor-outdoor*) ou vice-versa (*outdoor-indoor*), ambientes densamente urbanos, ambientes urbanos, ambientes suburbanos, ambientes residenciais e ambientes rurais [27] [24]. A figura 4 exemplifica um gráfico de nível de potência recebida, medida ao longo do tempo, com variação do afastamento entre transmissor e receptor. Nota-se grande variabilidade do sinal e queda com a distância, ocasionando o desvanecimento do mesmo ao longo do percurso de medição.

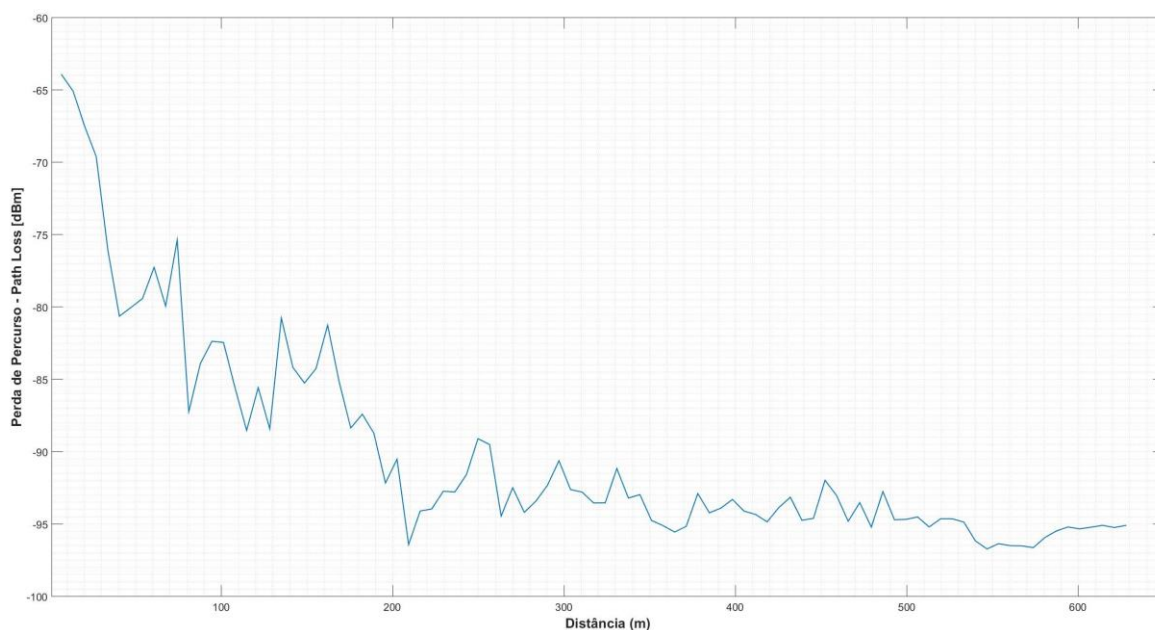


Figura 4- Variabilidade da potência do sinal ao longo do percurso de medição.

2.2.1 Desvanecimento (*fading*)

Em um CRM, à medida que o receptor se move pelo canal de propagação, os sinais recebidos por toda a extensão do percurso sofrem a influência dos mecanismos de propagação, de tal modo que ocorrem atenuações e constantes variações da amplitude do sinal recebido ao longo do tempo e da frequência, veja a figura 3 acima. Todas essas variações do sinal recebido correspondem ao fenômeno denominado desvanecimento (*fading*).

Steele [23] descreve uma ferramenta visualmente importante para projetar sistemas de comunicação rádio móvel. Tendo em conta todos os efeitos do desvanecimento (ganhos e atenuações de potência do sinal) sucedidos durante todo o percurso através do CRM e a potência de recepção necessária, é possível estimar a potência de transmissão necessária bem como a margem de potência para garantir a adequada e permanente comunicação entre origem e destino. Veja a figura 5 abaixo.

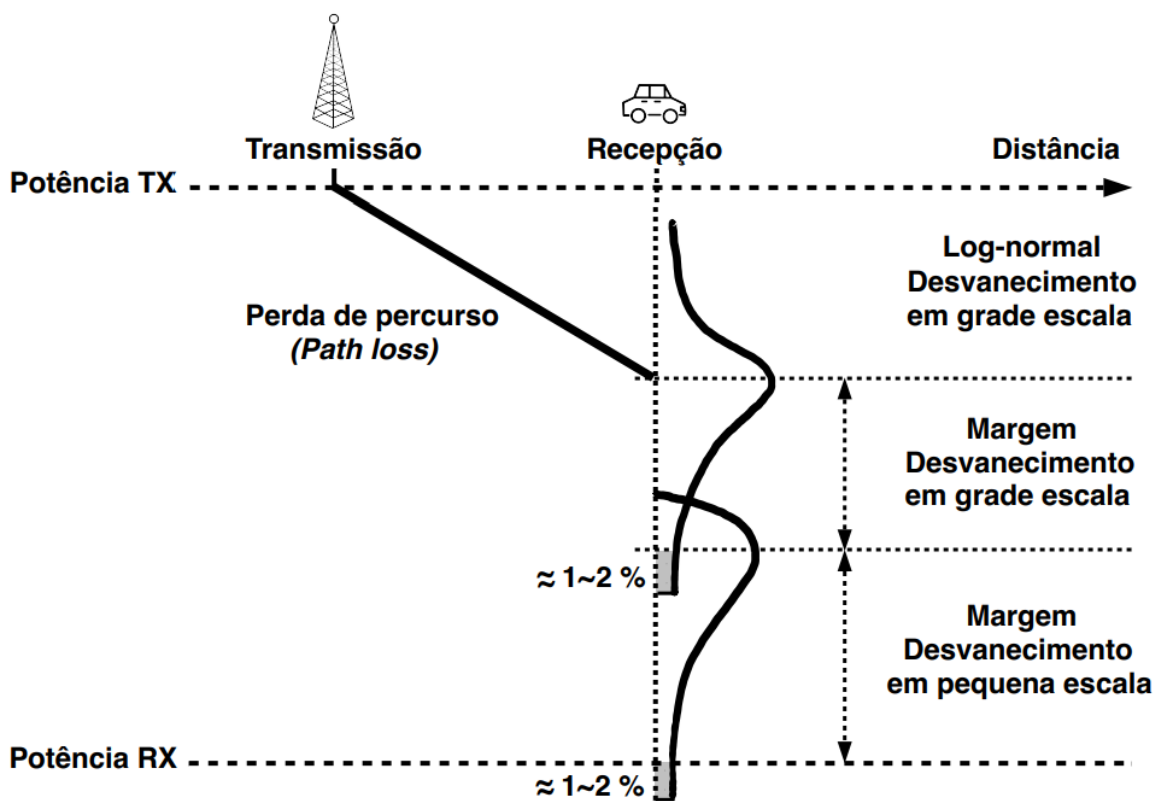


Figura 5- Ferramenta para projetos de sistemas de comunicação rádio móvel [23].

Devido à existência do desvanecimento em um CRM, a adequada caracterização desse canal ocorre segundo um processo dividido em duas etapas distintas: desvanecimento em grande escala e desvanecimento em pequena escala.

2.2.1.1 Desvanecimento em grande escala

O desvanecimento em grande escala (*large-scale fading*) [24], engloba a perda de percurso (*path loss*) e o sombreamento (*shadowing*) do sinal.

A perda de percurso corresponde a uma perda de natureza determinística que pode ser prevista como o decréscimo da potência do sinal recebido como função da distância.

O sombreamento diz respeito a uma perda de natureza aleatória, que corresponde à obstrução do sinal rádio móvel por grandes obstáculos, como prédios, montanhas e florestas (conforme a topografia local).

O sombreamento é adequadamente modelado por uma distribuição log-normal, com desvio padrão dependente da frequência e do canal de propagação [24]. A natureza log-normal indica que o sombreamento apresenta uma equiprobabilidade que varia em torno da perda média em cada ponto ao longo do percurso [28].

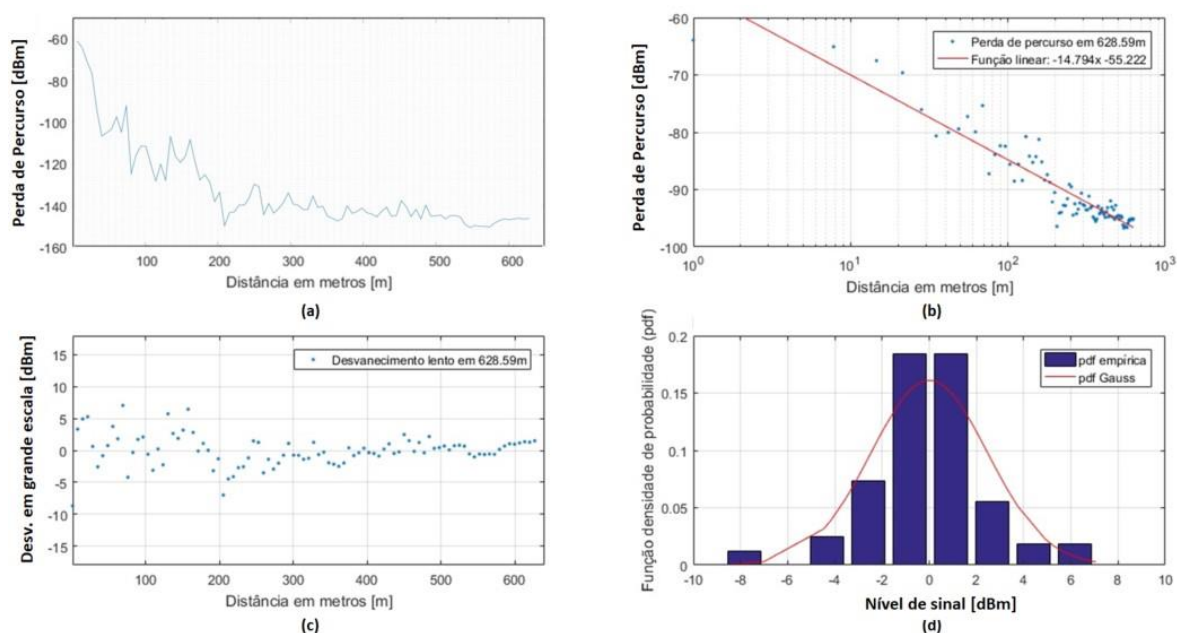


Figura 6- (a) Desvanecimento em grande escala ao longo do percurso. (b) Mesmo desvanecimento em escala logarítmica. Separação em perda de percurso e sombreamento. (c) Remoção da perda de percurso e visualização do efeito de sombreamento (d) Sombreamento apresenta natureza log-normal.

Como exemplo, a figura 6 acima expõe as várias componentes de uma medição faixa estreita realizada em ambiente *outdoor*. A figura 6 (a) apresenta a potência média recebida em cada ponto de medição ao longo do percurso. Essa potência sofre variações e se atenua no decorrer da trajetória, devido ao desvanecimento em grande escala. Colocando-se o eixo das abscissas em escala logarítmica e usando o método dos mínimos quadrados, figura 6 (b), pode-se traçar a perda de percurso como uma reta de melhor ajuste ao longo da trajetória. Percebe-se, assim, o comportamento determinístico de perda de

percurso e o comportamento aleatório que caracteriza o sombreamento. Na figura 6 (c), ao se subtrair a reta que representa a perda de percurso, observa-se apenas o sombreamento que apresenta forte correlação com as características geográficas do terreno. Na figura 6 (d), constata-se que o sombreamento apresenta um adequado ajuste à distribuição log-normal.

2.2.1.1 Desvanecimento em pequena escala

Ao contrário do desvanecimento em grande escala, que se preocupa com a atenuação média da potência do sinal ao longo de todo o percurso entre transmissor e receptor, o desvanecimento em pequena escala (*small-scale fading*) preocupa-se em investigar, individualmente, diferentes setores pontuais ao longo do percurso. A análise de cada setor, da ordem de meio comprimento de onda (λ) [24], busca compreender o somatório de efeitos pontuais causados pela propagação de múltiplos percursos (*multipath propagation*).

No CRM, os movimentos relativos entre transmissor e receptor causam deslocamentos na frequência do sinal recebido (efeito Doppler). Tais manifestações no domínio da frequência possuem efeitos duais no tempo, causando dispersões temporais. Ao se levar em conta a extensão dos múltiplos percursos, o desvanecimento em pequena escala pode interferir em todas as componentes de frequência do sinal de uma mesma maneira (*flat fading*) ou de maneiras distintas para diferentes componentes do sinal (*frequency-selective fading*). Pode-se, ainda, caracterizar, no domínio temporal, o desvanecimento em pequena escala como desvanecimento de variação de tempo (*time variance fading*).

Como exemplo, a figura 7 abaixo apresenta as várias componentes medidas em um ponto de medição (único setor). Enquanto a figura 6 (a) apresenta a potência média em cada setor, a figura 7 apresenta 8560 medições que compõem um único setor, caracterizando o desvanecimento em pequena escala. Pode-se perceber que as atenuações em torno de 15 dB são comuns e que existem atenuações mais severas ao longo do percurso, em torno de 40 dB. Todas essas atenuações e variações de potência dos sinais recebidos caracterizam a propagação pelos múltiplos percursos e ao conseqüente somatório, no receptor, das múltiplas ondas, cada uma apresentando diferentes amplitudes, fases e atrasos.

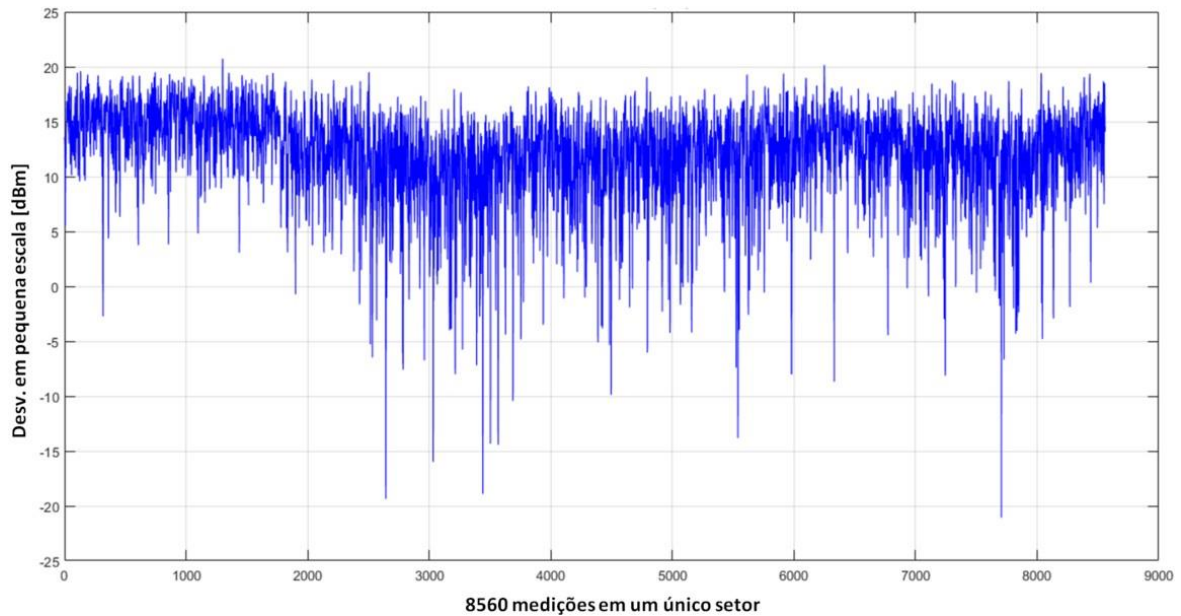


Figura 7- Desvanecimento em pequena escala em ambiente *outdoor*.

Visto que é melhor trabalhar com modelos, já que esses descrevem o comportamento de fenômenos como o desvanecimento em pequena escala, existem diversos modelos estatísticos que caracterizam tal desvanecimento e estudam a distribuição da envoltória do sinal recebido. Como exemplo, a figura 8 demonstra como os dados se encaixam nos modelos de distribuição normal, Rayleigh, Rice e m-Nakagami.

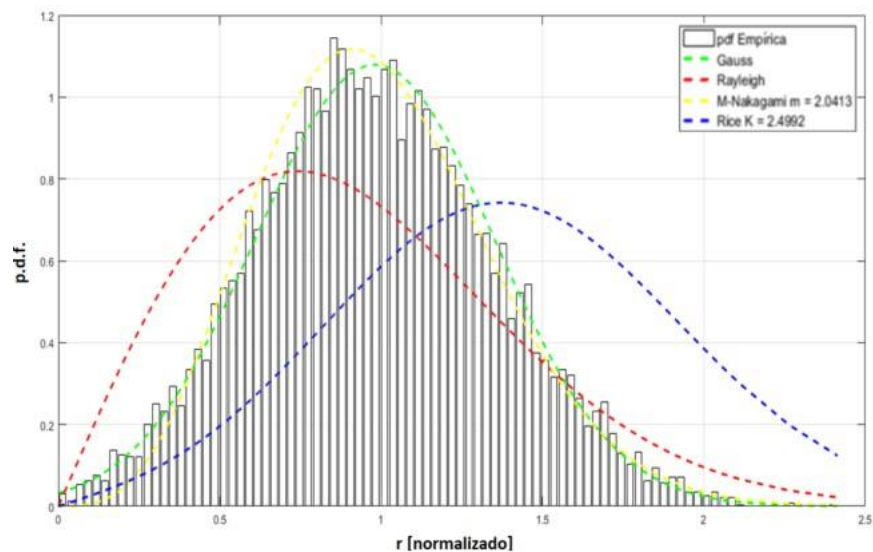


Figura 8- Potência do sinal medido em um ponto em ambiente interno (*indoor*).

2.3 Caracterização do canal rádio móvel

A caracterização e sondagem do canal rádio, especificado aqui como CRM, podem ser realizadas tanto para canais faixa estreita quanto para canais faixa larga e objetivam a análise estatística do canal rádio. Ou seja, os parâmetros adequados que caracterizam o CRM serão sondados, medidos e analisados, de maneira que se possa compreender corretamente o comportamento do canal por onde se propaga o sinal rádio.

2.3.1 Caracterização faixa estreita

Corresponde a transmissão de um sinal, representando por um tom. Nesse caso, a transformada de Fourier é impulsional na frequência do tom. Desse modo, a sondagem do canal rádio é realizada apenas nessa respectiva frequência (portadora). Para um CRM, como pode existir mobilidade relativa entre transmissor e receptor, haverá uma possível variação nas redondezas da frequência portadora, devido ao efeito Doppler.

Dessa forma, é possível analisar as variações do sinal devido às perdas de percurso e aos desvanecimentos em grande e em pequena escala, bem como detectar as funções e distribuições estatísticas que melhor se encaixam aos dados analisados.

Os exemplos apresentados nas figuras 6, 7 e 8 (tópicos 2.2.1.1. e 2.2.1.2), são boas representações da caracterização faixa estreita.

2.3.2 Caracterização faixa larga

Para a caracterização faixa larga do CRM, Parsons [24] considera, em um primeiro momento, o canal como um filtro linear com variação temporal determinística. Em um segundo momento, o canal é tratado como aleatoriamente variante no tempo, onde as respectivas funções passam a apresentar variáveis aleatórias. Por fim, o canal é simplificado para um modelo estacionário no sentido amplo com espalhadores decorrelacionados (WSSUS, *Wide-Sense Stationary Uncorrelated Scattering*).

Os conceitos acima descritos tomam como base uma sucessão de estudos. Zadeh [29] analisou filtros lineares variantes no tempo. Kailath [30] desenvolveu trabalho adicional com foco na caracterização do canal como um filtro linear aleatoriamente

variante no tempo, analisando modelos canônicos de amostragem de canais com variação temporal. Bello [31] [32] ampliou os trabalhos anteriores, demonstrando relações simétricas nos domínios da frequência e do tempo via dualidade e transformadas de Fourier, generalizando o conceito de dualidade tempo-frequência e apresentando um modelo estatístico generalizado para o canal rádio com base em suas funções representativas. Seguem abaixo o detalhamento sobre essas considerações.

2.3.2.1 Caracterização do canal rádio real

O objetivo aqui é desenvolver as ideias acima citadas, de maneira que as simplificações utilizadas para a análise de um canal rádio real sejam alcançadas.

Canais cujas estatísticas de variações temporais são consideradas constantes são denominados estacionários no sentido amplo (WSS, *Wide-Sense Stationary*). Se diferentes espalhadores contribuem com diferentes retardos na composição do sinal de resposta do canal rádio, eles são ditos descorrelacionados. Canais desse tipo são nomeados de canais com espalhadores descorrelacionados (US, *Uncorrelated Scattering*).

A caracterização do CRM em faixa larga é uma tarefa extremamente custosa, visto que o próprio canal não é estacionário, ele é aleatório. Nesse sentido, a correta caracterização estatística do canal exigiria o conhecimento das pertinentes funções de densidade de probabilidade conjunta multidimensionais das variáveis envolvidas no processo, as quais não estão disponíveis com as sondagens e medições oficiais realizadas.

Todavia, pode-se incorporar uma natureza de tempo invariante, ao acatar o processo de caracterização como um encadeamento de curtos espaços de tempo ou curtas distâncias de percurso, considerando, assim, que o canal possui o mesmo comportamento estatístico nesses pequenos intervalos [24]. Na verdade, parte-se da hipótese que um canal é WSS e US, ou seja, ele é estacionário no sentido amplo com espalhadores descorrelacionados (WSSUS, *Wide-Sense Stationary Uncorrelated Scattering*) e se conclui que essa suposição pode ser considerada apenas em distância pequenas ou intervalos temporais curtos. Em vista disso, o canal é considerado quase-estacionário no sentido amplo e com espalhadores descorrelacionados (QWSSUS, *Quase-Wide-Sense Stationary Uncorrelated Scattering*) e as suas funções de correlação podem ser simplificadas.

2.3.2.2 Técnicas de sondagem do canal banda larga

Dado a existência dos mecanismos de propagação e a influência desses mecanismos no canal de propagação, gerando múltiplos percursos que desencadeiam nos efeitos relacionados ao desvanecimento em pequena escala, várias técnicas de sondagem do canal banda larga foram desenvolvidas.

Cabe aqui, apenas citar algumas delas para ciência das mesmas. Logo, essas técnicas podem ser classificadas como [25]:

- Sondagem no domínio da frequência:
 - Medições de pulso direto (*direct pulse measurements*);
 - Medições de espectro espalhado com correlação deslizante (*spread spectrum sliding correlator measurements*);
 - Medições de varredura em frequência (*swept frequency measurements*).

- Sondagem no domínio do tempo:
 - Medições de pulso periódico;
 - Medições de filtro casado (LFM Chirp, sequência pseudo-aleatória – *PN sequence*);
 - Varredura de Retardo de tempo e correlação cruzada (*Swept-Time Delay Cross Correlation* – STDCC).

2.4 Parâmetros de dispersão do canal rádio móvel

Seguem os principais parâmetros de dispersão do CRM. Eles são muito úteis, pois fornecem importantes informações sobre para a adequada caracterização do canal rádio banda larga [24].

2.4.1 Parâmetros de dispersão no tempo

2.4.1.1 Retardo médio (*average delay*)

No contexto do canal rádio, o retardo médio (τ_m) representa a média dos tempos que múltiplos sinais levam para percorrer os múltiplos percursos até o receptor partindo do transmissor em um instante de tempo t . O retardo médio (τ_m) corresponde ao primeiro momento central ou estatística de primeira ordem (valor esperado) do perfil de potência de retardo (power-delay profile) $P_h(\tau)$. Parsons [24] define o retardo médio (τ_m) como:

$$\tau_m = \frac{\int_0^{\infty} \tau P_h(\tau) d\tau}{\int_0^{\infty} P_h(\tau) d\tau} \quad (2.1)$$

Uma vez que a técnica de varredura em frequência utiliza nesse presente trabalho, a equação (2.1) pode ser reescrita como um somatório de N picos existentes em cada perfil de potência de retardo $P_h(\tau)$.

$$\tau_m = \frac{\sum_{i=0}^{N-1} \tau_i P_h(\tau_i)}{\sum_{i=0}^{N-1} P_h(\tau_i)} \quad (2.2)$$

2.4.1.2 Espalhamento de retardo (*delay spread*)

No contexto do canal rádio, o espalhamento de retardo (σ_T), também chamado de retardo RMS, corresponde ao desvio padrão em relação a média dos tempos que múltiplos sinais levam para percorrer os múltiplos percursos até o receptor. O espalhamento de retardo (σ_T) representa a raiz quadrada do segundo momento central (variância) do o perfil de potência de retardo (power delay profile) $P_h(\tau)$.

$$\sigma_T = \sqrt{\frac{\int_0^{\infty} (\tau - \tau_m)^2 P_h(\tau) d\tau}{\int_0^{\infty} P_h(\tau) d\tau}} \quad (2.3)$$

Como descrito acima, a equação (2.3) pode ser reescrita como um somatório.

$$\sigma_T = \sqrt{\frac{\sum_{i=0}^{N-1} (\tau_i - \tau_m)^2 P_h(\tau_i)}{\sum_{i=0}^{N-1} P_h(\tau_i)}} \quad (3.4)$$

2.4.1.3 Banda de coerência (*coherence bandwidth*)

A banda de coerência (B_c) corresponde à medida estatística da largura de banda (faixa de frequência) sobre a qual a correlação entre as amplitudes das componentes espectrais é elevada, tipicamente maior que 90%. [25].

2.4.2 Parâmetros de dispersão na frequência

2.4.2.1 Deslocamento Doppler

O deslocamento Doppler, ou deslocamento médio, é representado por:

$$v_m = \frac{\int_0^{\infty} v P_H(v) dv}{\int_0^{\infty} P_H(v) dv} \quad (2.5)$$

Como descrito acima, a equação (2.5) pode ser reescrita como um somatório.

$$v_m = \frac{\sum_{i=0}^{N-1} v_i P_h(v_i)}{\sum_{i=0}^{N-1} P_h(v_i)} \quad (2.6)$$

2.4.2.2 Espalhamento Doppler (*Doppler spread*)

O espalhamento Doppler é representado por:

$$\sigma_D = \sqrt{\frac{\int_0^{\infty} (v - v_m)^2 P_H(v) dv}{\int_0^{\infty} P_H(v) dv}} \quad (2.7)$$

Como descrito acima, a equação (2.7) pode ser reescrita como um somatório.

$$\sigma_D = \sqrt{\frac{\sum_{i=0}^{N-1} (v_i - v_m)^2 P_h(v_i)}{\sum_{i=0}^{N-1} P_h(v_i)}} \quad (4.8)$$

2.4.2.3 Tempo de coerência (*coherence time*)

O Tempo de Coerência (T_c) é uma medida estatística que corresponde ao intervalo dentro do qual a resposta ao impulso do canal pode ser considerada aproximadamente invariante. Ou seja, existe forte correlação de amplitude entre dois sinais recebidos, correlação tipicamente maior que 90% [25].

2.5 Modelos de Markov e cadeias de Markov

O início do século XX foi inspirador para a teoria da probabilidade. Entre as contribuições para esse ramo da matemática, cabe destaque ao estudo inicialmente conduzido em 1906 por Andrei A. Markov sobre eventos conectados em cadeia. Desde meados do século passado, temas relacionados a esse tópico foram amplamente estudados, apresentando atualmente bases matemáticas e ferramentais completamente sólidas [33].

Em vários fenômenos físicos, químicos, sociais, biológicos, entre outros, percebem-se as aplicações dos estudos de Markov, visto que esses fenômenos podem ser modelados de tal forma que a probabilidade de evolução para um estado no instante seguinte esteja associada exclusivamente ao estado presente.

Cabe aqui, uma apresentação formal sobre o tema, já que o mesmo fundamenta métodos adotados no presente trabalho. Tal exposição objetiva uma definição inicial de termos e nomenclaturas para ambientação do leitor, de maneira que a base matemática adotada esteja sempre disponível para consultas e detalhamentos.

De fato, muitos processos estocásticos possuem estados que não são independentes entre si, podendo apresentar diferentes graus de dependência. Modelos ou processos markovianos são ditos como tipos especiais de processos estocásticos e, apesar de exibirem uma forma simples de dependência, possuem aplicabilidade na modelagem de diversos fenômenos. Logo, processos de Markov são processos estocásticos onde um estado

depende exclusivamente do seu estado antecessor e é completamente independente dos demais estados predecessores. Se os processos de Markov envolverem um espaço de estados discretos, numerável e finito, esses serão nomeados como cadeias de Markov. Em outras palavras, em uma cadeia de Markov, a partir de um estado, existe um conjunto definido, discreto e limitado de possíveis estados sucessores, cada um com sua respectiva probabilidade [26]. Dessa forma, a propriedade de Markov de primeira ordem especifica que:

$$\begin{aligned} P[Q(s_{t+1}) = q_{t+1} \mid Q(s_t) = q_t, Q(s_{t-1}) = q_{t-1}, \dots, Q(s_2) = q_2, Q(s_1) = q_1] \\ = P[Q(s_{t+1}) = q_{t+1} \mid Q(s_t) = q_t] \end{aligned} \quad (2.9)$$

Fala-se bastante em estados e, em uma cadeia de Markov, esses correspondem as possíveis opções de saídas dentro do conjunto discreto e limitado de opções do espaço amostral do experimento. Cabe destacar também que a passagem, ou o salto, de um estado para qualquer outro entre os estados disponíveis, será denominado de transição e, dentro do formalismo Markoviano, essa transição ocorre com certa probabilidade. Quando se assume que a probabilidade de transição de um estado para o estado subsequente seja estática e invariante no tempo, a cadeia de Markov é dita homogênea [34] [35]. Dessa maneira, independente do índice t , a probabilidade $P[Q(s_{t+1}) = q_{t+1} \mid Q(s_t) = q_t]$ torna-se fixa e pode ser escrita como $a_{ij} = P(q_{t+1} = s_j \mid q_t = s_i)$, onde se define a_{ij} como a transição homogênea e invariante no tempo do estado s_i para o estado s_j . Sendo a_{ij} , por assim dizer, chamada de função de transição.

$$a_{ij} = P(q_{t+1} = s_j \mid q_t = s_i) \quad \text{para todo } t \quad (2.10)$$

A função de transição a_{ij} pode ser representada matricialmente. Sendo ela homogênea, a matriz que a representa seguirá esse mesmo princípio atemporal e apresentará probabilidade de transição independente do tempo. Ela será uma matriz quadrada $N \times N$ (conforme o número de estados N), suas linhas apresentarão soma igual a um e ela será denominada matriz de probabilidade de transição de estados. Segue abaixo a representação da referida matriz:

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1N} \\ a_{21} & a_{22} & \cdots & a_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ a_{N1} & a_{N2} & \cdots & a_{NN} \end{bmatrix} \quad \text{onde } a_{ij} \geq 0 \quad \text{e} \quad \sum_{j=1}^N a_{ij} = 1 \quad (2.11)$$

Sendo assim, uma cadeia de Markov torna-se bem estruturada quando são conhecidos os parâmetros e variáveis envolvidos no processo, tais como:

1. Parâmetro t , que é uma representação temporal da sequência de estados, de tal forma que t representa o presente, $t+1$ representa o futuro e $1, \dots, t-1$, representam o passado.
2. Parâmetro N , que representa o número total de estados.
3. Conjunto discreto, numerável e finito de estados $S = \{s_1, s_2, s_3, \dots, s_N\}$.
4. Variável aleatória Q , que assinala um número real, associado à probabilidade de transição de estado, para cada possível estado s do conjunto discreto, numerável e finito de possíveis estados S .
5. Distribuição de probabilidade do estado inicial $R = \{r_i\}$ com $r_i = P(q_t = s_i)$ em $t = 1$. Logo, $r_i = P(q_1 = s_i)$, com $1 \leq i \leq N$.
6. Distribuição de probabilidade da transição de estados definida como $A = \{a_{ij}\}$, que corresponde a transição do estado s_i para o estado s_j . Tal que $a_{ij} = P(q_{t+1} = s_j | q_t = s_i)$, com $1 \leq i, j \leq N$.

As distribuições de probabilidade de transição de estados $A = \{a_{ij}\}$ e de probabilidade do estado inicial $R = \{r_i\}$ resumem as estruturas listadas acima e caracterizam completamente uma cadeia de Markov, de tal forma que é possível representar um modelo, nomeado aqui de M_{CM} , de forma compacta como $M_{CM} = (A, R)$ [35].

2.6 Modelos ocultos de Markov (HMM)

Visto que o tema modelos ocultos de Markov (HMM) embasou os métodos adotados no presente trabalho, uma apresentação formal torna-se adequada. Tal descrição objetiva a exposição e o detalhamento dos termos e nomenclaturas, para que toda a base matemática adotada seja adequadamente referenciada.

Embora as cadeias de Markov, expostas na seção anterior, apresentem grande aplicabilidade em diferentes contextos, nem todo fenômeno expõe os respectivos estados envolvidos de forma direta ao observador. Muitas vezes, as inferências sobre um fenômeno, cujos estados estão ocultos, pautam-se nas observações de incidentes secundários, que correspondem às ocorrências resultantes desse fenômeno escondido aos olhos do observador. Nesse contexto, uma possível modelagem envolve a adoção de duas camadas distintas, uma delas responsável por representar os estados ocultos e a outra responsável por retratar os incidentes secundários. Esses incidentes visíveis estão associados diretamente às ocorrências dos estados ocultos, ou estão associados às transições entre esses estados ocultos [34]. Essa adoção de uma dupla camada permite maior liberdade para a representação do fenômeno, enriquecendo sua modelagem e evitando entraves na estrutura básica das cadeias de Markov [34]. Portanto, pode-se dizer que um HMM representa um conjunto de dois processos estocásticos, sendo um deles invisível aos olhos do observador. Esse processo estocástico oculto é inferido e compreendido através da sequência de incidentes associados ao processo estocástico visível. Esses incidentes são aqui chamados de observáveis.

Cabe dizer que o processo estocástico oculto satisfaz a propriedade de Markov, de tal forma que um estado oculto depende exclusivamente do seu estado oculto antecessor e é completamente independente dos demais estados ocultos predecessores. Adicionado a isso, o processo estocástico observável é dependente da atividade do processo estocástico oculto, de tal forma que os resultados observáveis representam as saídas (associadas aos estados ocultos ou às transições entre estados ocultos) das atividades ocorridas no processo estocástico oculto [35].

Logo, por mais que estejam definidos os parâmetros do HMM, o processo que governa o sistema continua invisível aos olhos do observador, sendo possível apenas a observação indireta da evolução desse processo via o processo estocástico observável. Os observáveis, nesse caso, nada mais são que uma sequência de incidentes aleatórios que representam diretamente a evolução do processo estocástico visível e, indiretamente, a evolução do processo estocástico oculto. Essa representação ocorre ao longo de uma unidade definida, como o tempo, por exemplo. A figura 9 exemplifica o HMM.

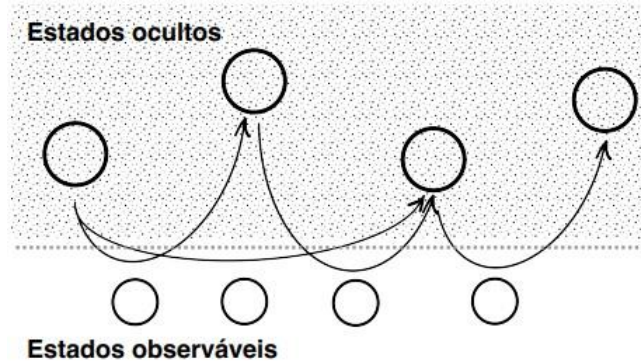


Figura 9- Representação do HMM.

Como feito para a cadeia de Markov, o HMM também se torna bem estruturado quando são conhecidos os parâmetros e variáveis desse método estatístico. Cabe ressaltar que além das definições presentes para a cadeia de Markov, são necessários novos parâmetros e variáveis para a composição dos observáveis, do processo estocástico observável e para diferenciação da dupla camada. Seguem as definições para o HMM:

1. Parâmetro t , que é uma representação temporal da sequência de estados. De tal forma que t representa o presente, $t + 1$ representa o futuro e $1, \dots, t - 1$, representam o passado.
2. Parâmetro N , que representa o número total de estados.
3. Conjunto discreto, numerável e finito de estados $S = \{s_1, s_2, s_3, \dots, s_N\}$.
4. A v.a. Q , que assinala um número real, associado à probabilidade de transição de estado, para cada possível estado s do conjunto discreto, numerável e finito de possíveis estados S .
5. Distribuição de probabilidade do estado inicial $R = \{r_i\}$ com $r_i = P(q_t = s_i)$ em $t = 1$. Logo, $r_i = P(q_1 = s_i)$, com $1 \leq i \leq N$.
6. Distribuição de probabilidade da transição de estados definida como $A = \{a_{ij}\}$, que corresponde a transição do estado s_i para o estado s_j . Tal que $a_{ij} = P(q_{t+1} = s_j | q_t = s_i)$, com $1 \leq i, j \leq N$.
7. Conjunto discreto, numerável e finito de observáveis $Y = \{y_1, y_2, y_3, \dots, y_M\}$. Dado que esses observáveis são gerados em cada estado s ou na transição do estado s_i para o estado s_j .
8. Parâmetro M , que representa o número total de observáveis.

9. A v.a. O , que assinala um número real, associado à probabilidade de transição de observáveis, para cada possível observável y do conjunto discreto, numerável e finito de possíveis observáveis Y .

Logo, $O(y_t) = o_t \rightarrow$ v.a. O está associada a uma saída/observável y do espaço de amostras de observáveis Y no índice t , assume um valor real o também no índice t .

10. Distribuição de probabilidade da transição de observáveis definida como $B = \{b_{ij}(k)\}$, associada ao observável produzido pela transição do estado s_i para o estado s_j . Tal que $b_{ij}(k) = P(o_t = y_k | q_{t-1} = s_i, q_t = s_j)$, com $1 \leq i, j \leq N$ e $1 \leq k \leq M$.
11. $k =$ Índice de inteiro que corresponde à transição entre observáveis. Ou corresponde ao tempo.

Logo, ao se usar as definições acima, é possível criar a sequência temporal de observáveis $O = o_1, o_2, \dots, o_t, \dots, o_T$ de tal forma que se define o seguinte algoritmo iterativo [17]:

1. Definir a representação temporal em $t = 1$ para início da modelagem com HMM.
2. Escolher o estado inicial $r_i = P(q_t = s_i)$ em $t = 1$ de acordo com a distribuição de probabilidade do estado inicial $R = \{r_i\}$.
3. Escolher o_t de acordo com a distribuição de probabilidade dos observáveis $b_{ij}(k) = P(o_t = y_k | q_{t-1} = s_i, q_t = s_j)$ associada à transição do estado s_i para o estado s_j .
4. Escolher o estado q_t de acordo com $a_{ij} = P(q_t = s_j | q_{t-1} = s_i)$ associada à transição do estado s_i para o estado s_j .
5. Definir $t = t + 1$. Retornar para o passo 3 se $t < T$. Caso $t = T$, terminar o algoritmo.

Como visto anteriormente, a estrutura $M_{CM} = (A, R)$ caracteriza de forma resumida uma cadeia de Markov. Da mesma maneira, para representar de forma compacta o HMM, torna-se necessário apenas adicionar a distribuição de probabilidade da transição de

observáveis $B = \{b_{ij}(k)\}$ a essa estrutura M_{CM} [16] [35], de tal forma que $M_{HMM} = (A, B, R)$.

2.7 Problemas controle do HMM

Como dito anteriormente, a criação de um modelo que descreva um fenômeno real resume-se na busca de uma representação matemática, mesmo que simplificada, do fenômeno em si. Para tanto, são necessárias as definições de parâmetros e variáveis usados para descrever o modelo estatístico, bem como de ajustes que permitirão o adequado enquadramento desse modelo ao fenômeno estudado [16]. Em outros termos, algoritmos iterativos com aprendizagem contínua são usados para estimar os parâmetros do modelo estatístico, de tal forma que a verossimilhança do conjunto de dados de treinamento tende a aumentar em cada iteração.

É importante dizer que a verossimilhança lida com ajustes do modelo em acordo com um conjunto de observações, conjunto de dados ou incidentes observados. Ao contrário da probabilidade que, perante uma distribuição de probabilidade, trabalha com a viabilidade de certos incidentes (medições observadas), $P(\text{incidentes} | \text{distribuição})$. Para a verossimilhança, dado um conjunto de incidentes, ajustam-se os parâmetros da distribuição de probabilidade para se adequarem aos dados observados, $P(\text{distribuição} | \text{incidentes})$. A maximização da verossimilhança diz respeito aos ajustes da distribuição de probabilidade de forma a maximizar a adequação dessa distribuição aos dados observados.

De modo prático e até computacional, é mais oportuno trabalhar com o logarítmico da verossimilhança. Pois, sendo o logarítmico uma função monótona, a maximização do logaritmo da função equivale à própria maximização da função em si. O uso do logarítmico simplifica a análise matemática e a precisão numérica computacional, visto que a soma de probabilidades logarítmicas é mais favorável que o produto de pequenas probabilidades [36] [37].

No que tange o HMM, cabem aqui três ajustes de tratabilidade e eficiência, chamados de problemas controle. Esses devem apresentar soluções para que a modelagem seja útil em aplicações reais.

2.7.1 Problema de avaliação (*evaluation problem*)

Determinar a probabilidade de uma dada sequência de observáveis $Y = \{y_1, y_2, y_3, \dots, y_M\}$ ser gerada por um modelo $M_{HMM} = (A, B, R)$ de forma computacionalmente eficiente, ou seja, um algoritmo para calcular $P(Y|M_{HMM})$.

Resposta para o problema de avaliação: Algoritmo *Forward-Backward*.

A probabilidade de Y dado o modelo M_{HMM} é chamado de verossimilhança. E o objetivo é maximizar essa probabilidade (função de verossimilhança).

2.7.2 Problema de decodificação (*decoding problem*)

Seja o modelo $M_{HMM} = (A, B, R)$ e a sequência de observáveis $Y = \{y_1, y_2, y_3, \dots, y_M\}$. Dentre as diversas sequências de estados que poderiam ter gerado essa sequência de observáveis, qual é a mais provável? Qual é a sequência ótima de estados?

Resposta para o problema de decodificação: algoritmo de Viterbi.

2.7.3 Problema de aprendizagem e treinamento (*learning problem*)

Como se ajusta os parâmetros do modelo $M_{HMM} = (A, B, R)$ para maximizar $P(Y|M_{HMM})$, probabilidade de uma sequência de observáveis?

Consiste em fazer o modelo $M_{HMM} = (A, B, R)$ se adaptar ao padrão da sequência de observáveis $Y = \{y_1, y_2, y_3, \dots, y_M\}$, chamada de sequência de treinamento (uma vez que é utilizada para treinar o HMM).

Resposta para Problema de aprendizagem e treinamento: Algoritmo de Baum-Welch, que usa o método conhecido como EM (*Expectation-Maximization*). Esse algoritmo é dividido em 2 etapas:

- Inicialização: definição das dimensões do modelo de Markov e valores iniciais das variáveis do HMM, ou seja, definição de: parâmetros fixos, valores iniciais dos parâmetros variáveis e da quantidade de iterações.

- Treinamento: o cálculo necessariamente dito. Onde os valores iniciais são re-estimados iterativamente (repetidamente).

2.8 Comparação da modelagem dos parâmetros do canal com a identificação de locutor através do uso de HMM

O reconhecimento da voz encontra variados empregos e, de forma geral, é possível listar aplicações como: permissão de acesso a áreas e a informações restritas, identificação via palavra-chave de um particular indivíduo ou de diferentes indivíduos, interação homem máquina para execução de atividades, etc.

O emprego do método HMM na modelagem estatística dos parâmetros do canal rádio foi inspirado no uso do HMM para identificação de um particular indivíduo, denominado de locutor. Para tal identificação, o conjunto de estados definidos para o HMM diz respeito aos fonemas relacionados à palavra-chave escolhida. Já para caracterização do canal rádio, o conjunto de estados diz respeito aos principais parâmetros de dispersão temporal do canal.

Dito isso, segue uma breve exposição da aplicação do HMM para identificação de locutor. Tal descrição objetiva não apenas a apresentação desse método, mas também busca facilitar a compreensão do uso do HMM para a modelagem do canal rádio, visto que será possível confrontar as ideias expostas em ambas as aplicações. Essa comparação torna-se útil em diversos momentos perante o emprego do HMM para a modelagem do canal rádio, uma vez que é mais fácil pensar em estados do HMM como fonemas de uma palavra-chave do que refletir sobre os mesmos estados como parâmetros de dispersão temporal do canal rádio.

Posto isso, é importante ressaltar que cada particular indivíduo apresenta características em seu aparelho fonador (conjunto de órgãos dos aparelhos digestivo e respiratório responsáveis pela fala) que o difere dos demais. De forma simplificada, o ar expelido pelos pulmões alcança a laringe, onde se modifica ao passar pelas pregas vocais para produzir sons, chamados vozeados ou sonoros. Caso as pregas vocais fiquem relaxadas, o ar passa sem produzir sons e são chamados de desvozeados ou surdos. Os sonoros, ao sair da laringe, são filtrados e amplificados e seguem em direção ao nariz (sons nasais) ou à boca (sons orais), onde são modificados de diferentes maneiras para

corresponderem aos sons desejados pelo locutor. Dessa forma, o corpo humano utiliza o conjunto de órgãos do aparelho fonador, mostrado na figura 10, para produzir sons, sendo os fonemas as menores unidades sonoras.

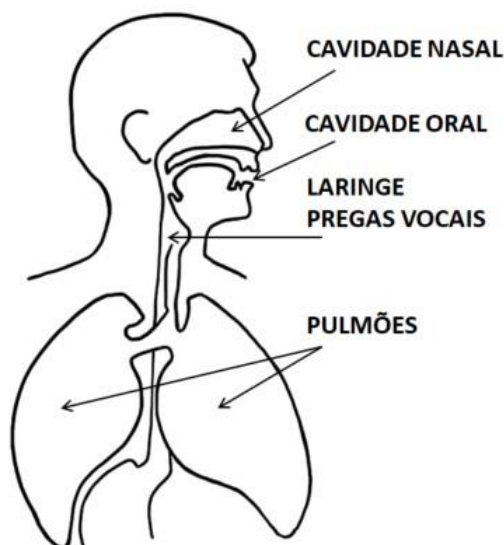


Figura 10- Aparelho fonador humano.

Assim sendo, cabe aqui uma descrição do processo que envolve o reconhecimento de palavra-chave de um particular indivíduo usando HMM. Posto que esse reconhecimento seja caracterizado por parâmetros que são particulares ao locutor em questão, tais sistemas individualizados precisam de uma fase de treinamento. Isso os difere, por exemplo, de sistemas independentes de locutor (onde o reconhecimento de palavra-chave independe do locutor), sendo esses sistemas já previamente treinados para reconhecer palavras-chaves de diferentes locutores. Logo, o primeiro passo para a identificação de locutor com uso de HMM consiste na definição de uma palavra-chave e gravação da mesma repetidas vezes, convertendo a energia acústica em energia elétrica e, posteriormente, em sinais digitais. Tal conjunto de sinais que caracterizam as gravações de pronúncias da palavra-chave forma a base de dados usada para treinamento e validação do sistema.

A aquisição da base de dados pode sofrer diversas influências do ambiente. Logo, torna-se necessário um pré-processamento para ajustes dos sinais. Esses podem apresentar componentes contínuas, as quais devem ser filtradas e removidas. Deve-se também normalizar as amplitudes, detectar e remover os períodos de silêncio no início e fim das pronúncias gravadas.

Após o pré-processamento, cada sinal passa por um filtro de pré-ênfase para destacar as componentes de alta frequência (importantes para a identificação do locutor), passa também por uma função janela para minimização de vazamentos espectrais. Em seguida, usa-se a transformada rápida de Fourier (FFT) para representação do sinal no domínio da frequência.

O espectro obtido é dividido por um banco de filtros passa-banda, cujas frequências centrais são linearmente espaçadas em uma escala denominada Mel e os envelopes são triangulares, para enfatizar as frequências centrais. Essa escala Mel, que faz referência à melodia, relaciona de forma logarítmica a frequência em Hertz com uma unidade de medida Mel. Tal unidade corresponde à percepção não linear do som pelo ouvido humano. É uma relação logarítmica entre a frequência em Hertz (f_{Hz}) e a frequência percebida pelo ouvido humano (f_{mel}), conforme figura 11 abaixo.

$$f_{mel} = 1127 \ln \left(1 + \frac{f_{Hz}}{700} \right) \quad (2.12)$$

Tal equação gera uma curva logarítmica, conforme esboçado na figura 11.

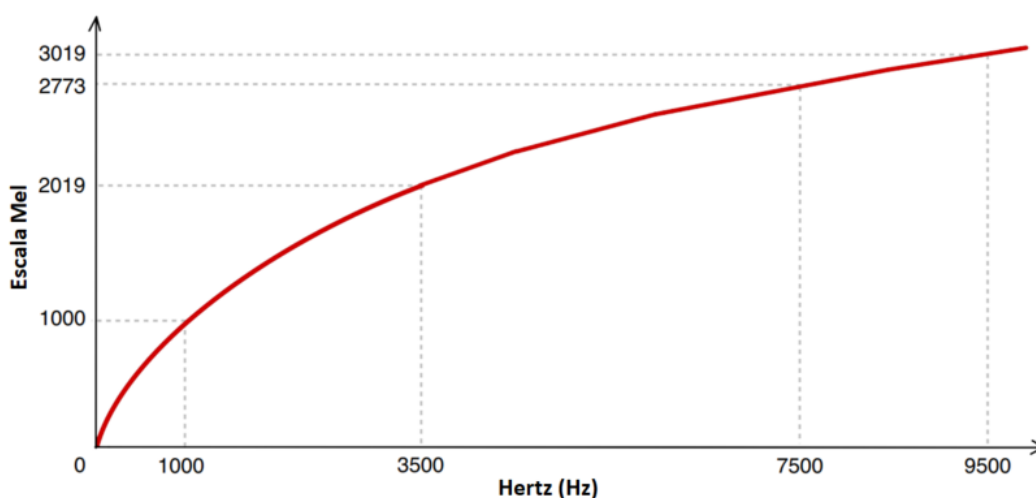


Figura 11- Escala Mel.

Após a adoção desse banco de filtros triangulares, linearmente espaçados de acordo com a escala Mel, aplica-se a Transformada Discreta de Cosseno e, dessa forma, geram-se os Coeficientes Cepstrais de Frequência Mel, denominados MFCCs, em referência ao termo em língua inglesa *Mel-Frequency Cepstral Coefficients*. Os MFCCs são usados para

extrair vetores de características de dados da voz, a fim de treinar um sistema de reconhecimento de palavra-chave de um particular indivíduo.

Conforme dito anteriormente, a inspiração para o emprego do método HMM na modelagem estatística do canal rádio surgiu do exposto acima. Então, da mesma forma que se buscou coeficientes MFCCs que caracterizassem a voz de um locutor, investigou-se parâmetros que descrevessem o canal rádio em um particular ambiente. Para tanto, o conjunto de dados do canal rádio precisou também ser adquirido, pré-processado e, posteriormente, tratado para que se alcançassem os vetores de características do canal rádio, aqui tratados como os principais parâmetros de dispersão temporal do canal: retardo médio (τ_m) e espalhamento de retardo (σ_T).

2.9 Topologias para transição de estados

Ao se aplicar o método estatístico HMM na modelagem e caracterização do canal rádio, certas questões revelaram-se importantes. Quando se pensa em um sistema que pode ser descrito como um conjunto de estados e que as passagens entre estados correspondem às transições, talvez, em um primeiro momento, se imagine apenas um conjunto ordenado de estados completamente conectados por transições sucessivas. Embora isso pareça razoável, existem variadas configurações que podem ser adotadas, existem variados caminhos, diferentes transições entre estados que podem ser legitimadas. A análise dessas configurações, descritas aqui como topologias, e a posterior escolha de uma opção entre as opções existentes, definiram como o sistema adotado deveria se comportar no modelo em questão.

A reflexão sobre a escolha de uma topologia propícia para o uso do HMM na modelagem de sistemas reais é de grande valia. Isso fica claro quando se pensa nas iterações necessárias ao sistema e na quantidade de processamento para treinamento do modelo. Não é difícil imaginar que os diferentes caminhos que levem a um determinado estado aumentem exponencialmente com mais transições. Melhor dizendo, quanto mais livre a topologia, quanto maior o número de possíveis transições, maior será o número de possíveis caminhos entre estados a serem percorridos [15] [16].

Como parte do estudo, seguem abaixo algumas topologias. Veja que existem variadas configurações que podem se adotadas, onde as transições podem percorrer os caminhos entre estados de diferentes maneiras, mantendo ainda certo formalismo.

Considerando a seguinte matriz de probabilidade de transição de estados $A = \{a_{ij}\}$, a função de transição a_{ij} pode ser representada matricialmente e, dessa forma, cabem as descrições de diferentes e possíveis topologias para implementação do HMM.

2.9.1 Topologia ergódica (*ergodic*)

A topologia ergódica, representada na figura 12, é uma topologia completamente conectada e sem restrições, onde cada estado pode alcançar qualquer outro dos possíveis estados em uma única transição.

Deste modo, essa topologia apresenta uma estrutura onde todos os estados estão conectados entre si, ou seja, o modelo apresenta uma matriz de probabilidade de transição de estados A onde, para todo i e j , a probabilidade $a_{ij} = P(q_t = s_j | q_{t-1} = s_i)$ será maior que zero.

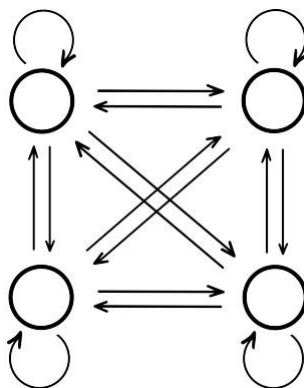


Figura 12- Topologia ergódica.

Segue a representação da matriz de probabilidade de transição de estados para a topologia ergódica.

$$A_{ergódica} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1N} \\ a_{21} & a_{22} & \cdots & a_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ a_{N1} & a_{N2} & \cdots & a_{NN} \end{bmatrix}, \text{ onde } a_{ij} > 0 \text{ para todo } i \text{ e } j \quad (2.13)$$

2.9.2 Topologia esquerda-direita (*left-to-right*)

Tomando como base a topologia ergódica, restringem-se agora as possíveis transições de um estado apenas para o seu respectivo conjunto de estados futuros. Em outras palavras, o estado presente nunca será referenciado por um índice maior que o índice do estado futuro. Veja sua representação na figura 13.

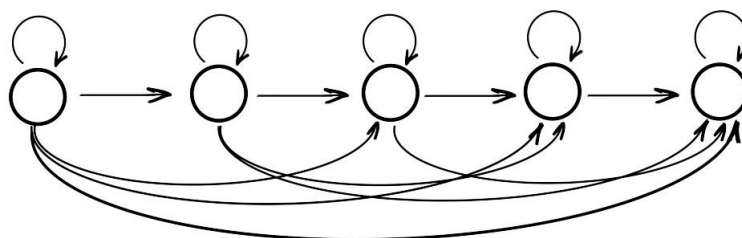


Figura 13- Topologia esquerda-direita.

Veja que, desse modo, a topologia apresenta uma estrutura temporal e se ajusta bem a modelos cujas propriedades variam com o tempo. Assim sendo, com o incremento temporal, o estado presente permanece o mesmo ou transiciona para um dos possíveis estados posteriores a ele. Dessa forma, a matriz de probabilidade de transição de estados A aponta para uma matriz triangular superior onde $a_{ij} \neq 0$ para $i \leq j$ e $a_{ij} = 0$ para $i > j$.

Segue a representação da matriz de probabilidade de transição de estados para a topologia esquerda-direita.

$$A_{\text{esquerda-direita}} = \begin{bmatrix} a_{11} & a_{12} & a_{13} & \cdots & a_{1N} \\ 0 & a_{22} & a_{23} & \cdots & a_{2N} \\ 0 & 0 & a_{33} & \cdots & a_{3N} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & a_{NN} \end{bmatrix} \quad (2.14)$$

2.9.3 Topologia de Bakis

Tomando como base a topologia esquerda-direita, em vista ainda de maior harmonia entre flexibilidade e eficiência, a topologia de Bakis restringe um pouco mais as possíveis transições entre estados. Além de se ater às transições para estados futuros,

permitem-se apenas transições para estados não tão distantes do estado atual. Na figura 14 está representada uma topologia de Bakis.

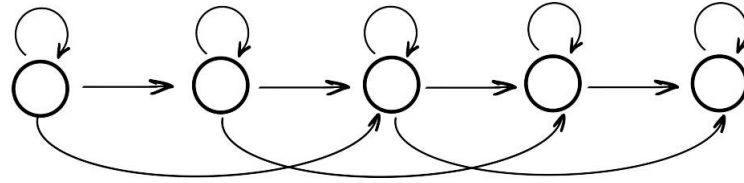


Figura 14- Topologia de Bakis.

Por assim dizer, essa topologia permite a permanência no estado atual, a transição para o estado seguinte ou apenas o salto do estado subsequente. Logo, ela ainda apresenta uma estrutura temporal e se ajusta bem aos modelos cujas propriedades variam com o tempo. A matriz de probabilidade de transição de estados A exibe valores não nulos apenas para a_{ij} onde $j = i$, $j = i + 1$ ou $j = i + 2$.

Segue a representação da matriz de probabilidade de transição de estados para a topologia de Bakis.

$$A_{Bakis} = \begin{bmatrix} a_{11} & a_{12} & a_{13} & 0 & \dots & 0 \\ 0 & a_{22} & a_{23} & a_{24} & \dots & 0 \\ 0 & 0 & a_{33} & a_{34} & \dots & 0 \\ 0 & 0 & 0 & a_{44} & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & 0 & \dots & a_{NN} \end{bmatrix} \quad (2.15)$$

2.9.4 Topologia linear

Tomando como base a topologia esquerda-direita, priorizando agora eficiência em contraste com a flexibilidade, a topologia linear apresenta-se de modo mais simples, de tal forma que ela permite a permanência no estado atual ou a transição para o estado subsequente, como se vê na figura 15.

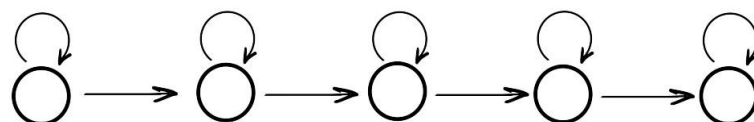


Figura 15- Topologia linear.

Essa topologia apresenta-se mais restritiva e, desse modo, a probabilidade $a_{ij} = P(q_t = s_j | q_{t-1} = s_i)$ será maior que zero apenas quando a transição persistir no mesmo estado ($j = i$) ou quando transicionar para o estado posterior ($j = i + 1$). Caso contrário, a_{ij} será igual a zero.

Segue a representação da matriz de probabilidade de transição de estados para a topologia linear.

$$A_{linear} = \begin{bmatrix} a_{11} & a_{12} & 0 & 0 & \dots & 0 \\ 0 & a_{22} & a_{23} & 0 & \dots & 0 \\ 0 & 0 & a_{33} & a_{34} & \dots & 0 \\ 0 & 0 & 0 & a_{44} & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & 0 & \dots & a_{NN} \end{bmatrix} \quad (2.16)$$

Capítulo 3 - Sondagem do canal rádio e cenários de medições

Diferentes técnicas podem ser empregadas para se caracterizar o canal rádio. A escolha adequada de uma determinada técnica deve levar em conta o tipo de canal utilizado, a aplicação prevista para o próprio canal, bem como a disponibilidade de equipamentos de medição apropriados. Ao se escolher adequadamente a técnica, facilita-se a caracterização desejada do canal rádio.

3.1 Adoção da técnica de varredura em frequência

A técnica utilizada na sondagem do canal rádio, dada a banda larga em análise, foi a técnica de sondagem no domínio da frequência, também chamada de varredura em frequência (*swept frequency – chirp*).

A viabilidade de medição do canal rádio usando uma técnica de resposta do canal ao impulso no domínio da frequência torna-se exequível para o presente estudo, uma vez que existe a relação dual entre os domínios do tempo e da frequência. Rappaport [25] argumenta que a técnica funciona bem, embora alegue a existência de moderações como calibração adequada e apurado sincronismo entre transmissor e receptor. Essas moderações delimitam as medições para ambientes de curta distância, em geral, ambientes *indoor*. Além disso, para canais que variam com o tempo, a resposta em frequência pode variar com grande rapidez, resultando em medições equivocadas de resposta ao impulso. Esse último obstáculo pode ser contornado com o uso de intervalos rápidos de varredura, diminuindo ao máximo o tempo de varredura. Um intervalo rápido de varredura pode ser obtido diminuindo o número de portadoras transmitidas, mas isso impacta a resolução do canal e o intervalo de atraso excessivo no domínio do tempo.

Em vista das moderações listadas acima, percebe-se que existe um limiar para o adequado uso da técnica de varredura em frequência. A campanha de medições e coleta oficial dos dados para o presente estudo tomaram os devidos cuidados com as calibrações e sincronismos, bem como com a distância entre transmissor e receber e o ambiente, garantindo na medida do possível uma estacionariedade do canal, visto que os intervalos de

varredura foram rápidos (com transmissão adequada de sinais) e as medições foram realizadas em períodos de baixo movimento.

3.2 Campanha de medições

A campanha de medições foi concentrada nos Campi da Praia Vermelha e do Gragoatá, localizados na cidade de Niterói, Rio de Janeiro, Brasil [38]. Ela engloba os processos de definição e preparação dos ambientes de medições, os processos relacionados à montagem, calibração e teste dos equipamentos utilizados, bem como as medições oficiais e definitivas.

3.2.1 Ambientes da campanha de medições

Os ambientes definidos para a campanha de medições foram escolhidos previamente. De forma geral, eles correspondem aos ambientes ditos interno (*indoor*), externo (*outdoor*), ambiente de transição de um local interno para um externo (*indoor-outdoor*), ou ambiente de transição de um local externo para um interno (*outdoor-indoor*).

Adicional cautela foi adotada para que as medições não sofressem grandes alterações e representassem efetivamente apenas o comportamento do canal rádio durante o tempo de propagação do sinal rádio transmitido. Uma vez que o canal apresenta-se como um fenômeno dinâmico, associando efetivamente sinais de excitação e resposta através de processos estocásticos, quaisquer alterações no ambiente, durante cada varredura de sinal, revelariam mudanças nas propagações dos sinais. Posto isso, as medições foram realizadas em períodos noturnos e nos finais de semana, evitando-se a alteração dos dados coletados pelo simples deslocamento de pessoas e objetos ao redor dos ambientes de medições, durante cada varredura de sinal. Os próprios pesquisadores tomaram cuidado para não obstruírem a linha de visada direta entre as antenas e se mantiveram imóveis durante o período de medições para garantir, da forma mais real possível, a estacionariedade do ambiente e, necessariamente, do canal rádio pesquisado a durante cada varredura de sinal.

As medições foram realizadas em diferentes ambientes que configuram diferentes cenários como: *indoor-outdoor*, *outdoor-indoor*, ambientes *outdoor* próximos a construções, ambientes *outdoor* amplos, mas não totalmente abertos e ambientes

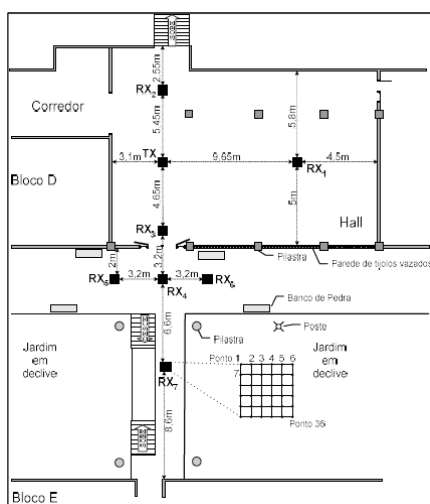
totalmente abertos (*outdoor*) [38]. Nos ambientes totalmente abertos, as medições foram realizadas em linha reta. Nos outros cenários as medidas foram realizadas utilizando-se uma malha quadrangular 6 por 6, com 36 posições.

Para essa dissertação, foram usados 8 diferentes ambientes:

- A5 (*indoor-outdoor*);
- A6 (*outdoor-indoor*);
- A7 (*indoor-outdoor*);
- A8 (*outdoor-indoor*);
- B3 (*outdoor*);
- B4 (*outdoor*);
- C1 (*outdoor*);
- C4 (*outdoor*).

3.2.1.1 Ambiente de medição A5 (*indoor-outdoor*)

O ambiente A5 corresponde ao *hall* do prédio de engenharia da UFF, bloco D, campus da Praia Vermelha. Medições realizadas com transmissor em ambiente *indoor*. Como o receptor movimentou-se por diferentes ambientes *indoor* e *outdoor*, esse ambiente ficou definido como ambiente *indoor-outdoor*. O ambiente A5 pode ser visto na figura 16.



(a)

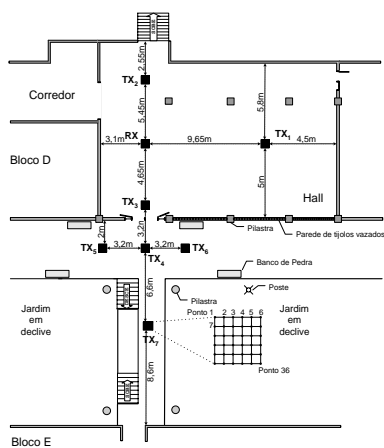


(b)

Figura 16- (a) Planta do ambiente A5. (b) Vista do ambiente A5.

3.2.1.2 Ambiente de medição A6 (*outdoor-indoor*)

O ambiente A6 corresponde ao *hall* do prédio de engenharia da UFF, bloco D, campus da Praia Vermelha. Medições realizadas com transmissor em ambiente *outdoor*. Como o receptor movimentou-se por diferentes ambientes *indoor* e *outdoor*, esse ambiente ficou definido como ambiente *outdoor-indoor*. O ambiente A6 pode ser visto na figura 17.



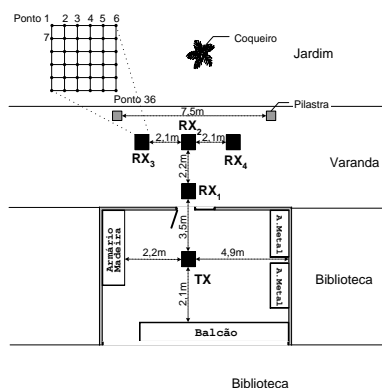
(a)



(b)

Figura 17- (a) Planta do ambiente A6. (b) Vista do ambiente A6.

3.2.1.3 Ambiente de medição A7 (*indoor-outdoor*)



(a)



(b)

Figura 18- (a) Planta do ambiente A7. (b) Vista do ambiente A7.

O ambiente A7 corresponde ao ambiente em frente à biblioteca de engenharia da UFF, Campus da Praia Vermelha. Medições realizadas com transmissor em ambiente

indoor. Como o receptor movimentou-se por diferentes ambientes *indoor* e *outdoor*, esse ambiente ficou definido como ambiente *outdoor-indoor*. O ambiente A7 pode ser visto na figura 18.

3.2.1.4 Ambiente de medição A8 (*outdoor-indoor*)

O ambiente A8 corresponde a uma passarela que liga os blocos D e E, localizada no terceiro andar do prédio de engenharia da UFF, campus da Praia Vermelha. As medições foram realizadas com transmissor em ambiente *outdoor*. Como o receptor movimentou-se por diferentes ambientes *indoor* e *outdoor*, esse ambiente ficou definido como ambiente *outdoor-indoor*. Cabe destacar que as localidades *outdoor* possuem visada e a localidade *indoor* não possui (atravessando a parede). O ambiente A8 pode ser visto na figura 19.



Figura 19- (a) Planta do ambiente A8. (b) Vista do ambiente A8.

3.2.1.5 Ambiente de medição B3 (*outdoor*)

O ambiente B3 corresponde ao prédio da faculdade de Geociências da UFF, bloco O, Campus da Praia Vermelha. Medições realizadas com transmissor e receptor em ambiente *outdoor*, o que definiu o ambiente de medição B3 como ambiente *outdoor*. O ambiente B3 pode ser visto na figura 20.

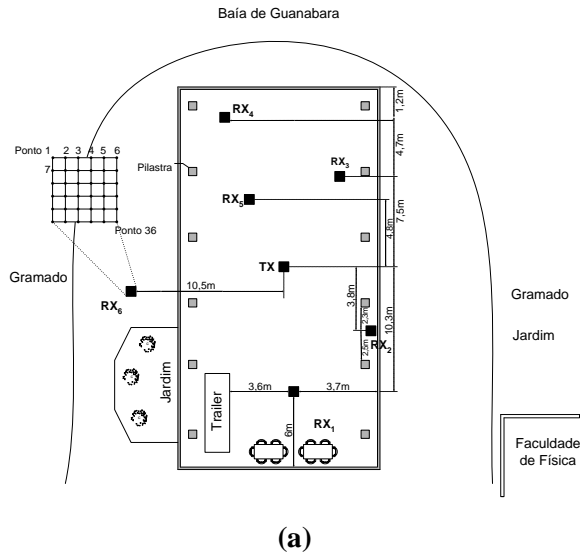


Figura 20- (a) Planta do ambiente B3. (b) Vista do ambiente B3.

3.2.1.6 Ambiente de medição B4 (*outdoor*)

O ambiente B4 corresponde a uma área de estacionamento localizada ao lado do bloco E, campus da Praia Vermelha. As medições foram realizadas com transmissor e receptor em ambiente *outdoor*, o que definiu o ambiente de medição B4 como ambiente *outdoor*. O ambiente B4 pode ser visto na figura 21.

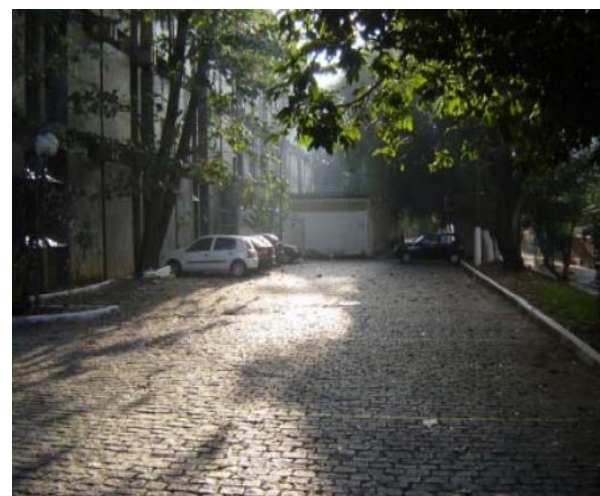
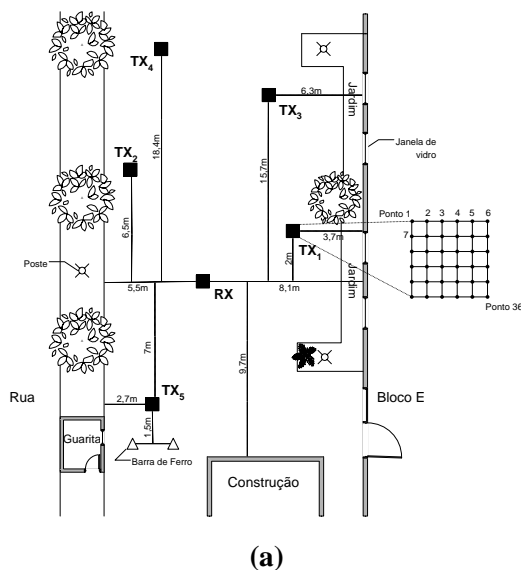
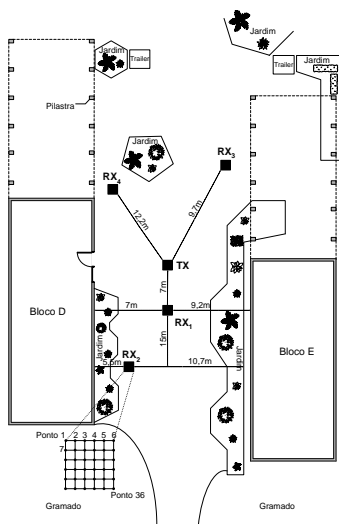


Figura 21- (a) Planta do ambiente B4. (b) Vista do ambiente B4.

3.2.1.7 Ambiente de medição C1 (*outdoor*)

O ambiente C1 corresponde a um espaço localizado entre os blocos D e E do campus do Gragoatá. Medições realizadas com transmissor e receptor em ambiente *outdoor*, o que definiu o ambiente de medição C1 como ambiente *outdoor*. O ambiente C1 pode ser visto na figura 22.



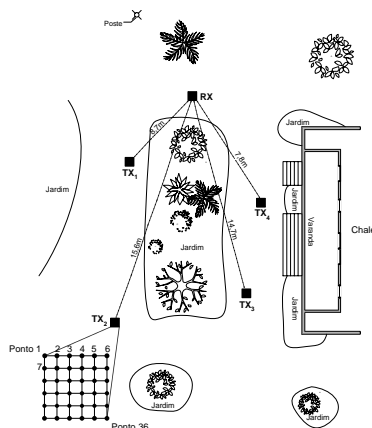
(a)



(b)

Figura 22- (a) Planta do ambiente C1. (b) Vista do ambiente C1.

3.2.1.8 Ambiente de medição C4 (*outdoor*)



(a)



(b)

Figura 23- (a) Planta do ambiente C4. (b) Vista do ambiente C4.

O ambiente C4 corresponde a uma área *outdoor* localizada no pátio em frente a casa de arquitetura, ao lado do casarão, no campus da Praia Vermelha. As medições foram realizadas com transmissor em ambiente *outdoor*. Medições realizadas com transmissor e receptor em ambiente *outdoor*, o que definiu o ambiente de medição C4 como ambiente *outdoor*. O ambiente C4 pode ser visto na figura 23.

3.2.2 Equipamentos da campanha de medições

A campanha de medições utilizou os seguintes equipamentos:

- Analisador de rede vetorial (modelo Hewlett Packard HP8714ET).
- Amplificador de baixo ruído (*Low Noise Amplifier* - LNA, modelo MINI-Circuits ZLR-2150) com faixa de aplicação de 950 MHz a 2150 MHz, ganho de 25 dB na faixa de 950 Mhz a 1800 MHz.
- Fonte de tensão de 12VCC que serviu de alimentação contínua para o LNA.
- Antenas banda larga do tipo discônica, omnidirecionais no plano horizontal que, na recepção, permitiu a captação de réplicas do sinal provenientes de todas as direções. As antenas de transmissão e recepção foram montadas em alturas iguais de 1,5 m do chão.
- Computador para armazenamento das medições e controle da sondagem do canal (controle do processo de varredura do canal).
- Cabo HPIB (*Hewlett Packard Interface Bus*) e placa de aquisição específica HP (Hewlett Packard) para conexão do computador com o analisador de rede vetorial. O protocolo de comunicação foi desenvolvido via programação em Matlab e permitiu a adequada comunicação entre o computador e o analisador de rede vetorial.
- Cabos coaxiais, modelo RG-213 e RG-58U.
- Conectores tipo N e tipo SMA.

3.2.3 Preparação (*setup*) dos equipamentos e ambientes

O analisador de rede vetorial atuou como um transceptor, de tal forma que gerou e transmitiu, por uma das portas, repetidos sinais senoidais de 10 dBm igualmente espaçados. Esses sinais senoidais, transmitidos sucessivamente, percorreram toda a banda de interesse, indo de 960 MHz até 1710 MHz. Por outra porta, o analisador de rede vetorial recebeu os sucessivos sinais transmitidos ao longo de toda a banda.

Para tanto, antes do início das medições, foi necessário montar os equipamentos, calibrar e testar os mesmos para que o analisador de rede vetorial funcionasse adequadamente, confrontando os sinais transmitidos por uma porta com os sinais recebidos por outra porta. Para que essa comparação entre sinais fosse adequada e refletisse apenas o comportamento do canal rádio, apenas o comportamento do ambiente por onde os sinais propagaram até chegar ao receptor, os efeitos e as perdas (perda de potência, de propagação, ruído) dos equipamentos envolvidos foram analisados e mitigados. Esse ajuste foi efetuado via função *Enhanced Calibration*, através da qual o analisador de rede vetorial compensou as perdas dos equipamentos utilizados, incorporando seus efeitos indesejados.

A inspeção e calibração do conjunto de equipamentos foram tais que os efeitos não influenciaram as coletas dos dados desejados durante as fases de medições oficiais. Para esse fim, a calibração deveria ser efetuada via ligação direta entre todos os equipamentos utilizados, o que indicaria os efeitos e as perdas de tais equipamentos. Como essa ligação direta não era possível em um conjunto que utilizasse antenas, a ligação direta entre as portas transmissora e receptora foi realizada diretamente via cabos e conectores, desprezando as respectivas antenas. Tal calibração aderiu então às perdas de propagação dos sinais nos respectivos cabos e conectores.

Após calibração via ligação direta com cabos e conectores, adicionando o conjunto de antenas transmissora e receptora, perceberam-se atrasos adicionais, de aproximadamente 2 ns, em todas as medições realizadas. Atrasos esses que diziam respeito às propagações dos sinais através das respectivas antenas transmissora e receptora, uma vez que esses atrasos não foram identificados anteriormente durante o processo de calibração com ligação direta. Tais atrasos também foram compensados para não influenciarem as medições oficiais.

Com os devidos ajustes e calibrações, definiram-se os seguintes parâmetros:

- A faixa de frequência sondada, de 960 MHz até 1710 MHz, corresponde à banda de frequência B em análise, sendo B igual a 750 MHz.
- O tempo de coleta de dados do analisador de rede vetorial foi de 696 ms. O tempo de gravação dos dados coletados foi de 304 ms. Portanto, a cada segundo (696 ms + 304 ms), o analisador de rede vetorial colheu (em 696 ms) e salvou (em 304 ms) os sinais transmitidos pelo canal rádio durante as fases oficiais de medições. O tempo de aquisição de 696 ms correspondeu ao menor período suportado pelo equipamento para percorrer as componentes espectrais geradas ao longo da banda de frequência analisada.
- O analisador de rede vetorial originou 1601 componentes espectrais espaçadas igualmente entre si. Esse espaçamento em frequência, nomeado de Δf , é aproximadamente igual a 0,47 MHz.

$$\Delta f = \frac{B}{(n^{\circ} \text{ pontos} - 1)} = \frac{750}{(1601 - 1)} = 0,46875 \text{ MHz} \cong 0,47 \text{ MHz} \quad (3.1)$$

- Outro parâmetro está associado ao somatório dos multipercursos entre os pontos de transmissão e recepção e diz respeito ao maior valor de retardo que se consegue medir. Esse valor, nomeado de Retardo Máximo ($\tau_{\text{máx}}$), é aproximadamente igual a $2,1333 \times 10^3$ ns.

$$\tau_{\text{máx}} = \frac{1}{\Delta f} = \frac{1}{0,47 \text{ MHz}} = 2,1333 \times 10^{-6} \text{ s} = 2,1333 \times 10^3 \text{ ns} \quad (3.2)$$

À vista disso, $\tau_{\text{máx}}$ permite afirmar que percursos com até 640 m podem ser detectados.

$$c = \lambda f = \frac{\lambda}{T} \rightarrow \lambda = cT = 3 \times 10^8 \times 2,1333 \times 10^3 \times 10^{-9} \cong 640 \text{ m} \quad (3.3)$$

- Um quarto parâmetro está associado à mínima diferença temporal entre a aquisição de dois raios, de tal forma que esses raios sejam distinguidos pelo

analisador de rede vetorial. Esse parâmetro, nomeado de Resolução Temporal ($\Delta\tau$), é igual a 1,333 ns.

$$\Delta\tau = \frac{1 \times 10^3}{B \times 10^3} = 1,333 \text{ ns} \quad (3.4)$$

Deste modo, essa diferença mínima no tempo possibilita calcular a menor distância percorrida por dois raios, para a qual o analisador de rede vetorial ainda consegue distingui-los. Componentes de multipercursos separados de até 40 cm podem ser distinguidos pelo analisador de rede vetorial, na sondagem efetuada.

$$c = \lambda f = \frac{\lambda}{T} \rightarrow \lambda = cT = 3 \times 10^8 \times 1,333 \times 10^{-9} \cong 40 \text{ cm} \quad (3.5)$$

A resolução temporal $\Delta\tau$ e o retardo máximo $\tau_{\text{máx}}$ são resultantes das configurações adotadas no analisador de rede vetorial durante toda a campanha de medições.

Tabela 1 - Parâmetros do sistema de sondagem.

Parâmetro	Valor	Unidade
Faixa de Frequência	960 a 1710	MHz
Largura de Banda	750	MHz
Potência de Transmissão	+10	dBm
Ganho do LNA	25	dB
Ganho das Antenas	2,14	dBi
Espaçamento em frequência entre amostras - Δf	0,46875	MHz
Número de amostras varridas	1601	-
Tempo de varredura	696	ms
Resolução Temporal - $\Delta\tau$	1,333	ns
Retardo Máximo - $\tau_{\text{máx}}$	2133	ns

- O uso de malhas quadrangulares de 36 posições permitiu o levantamento das estatísticas de pequena escala, visto que em cada ponto da malha o sinal foi recebido de forma descorrelacionada dos demais pontos. Parsons [24], a distância de $0,38\lambda$ representa o primeiro zero de correlação. A faixa de frequência é larga, então λ varia consideravelmente entre a frequência mais baixa λ_1 e a mais alta λ_2 .

$$\lambda_1 = \frac{3 \times 10^8}{960 \times 10^6} = 0,3125 \rightarrow 0,38 \lambda_1 \cong 11,875 \text{cm} \quad (3.6)$$

$$\lambda_2 = \frac{3 \times 10^8}{1710 \times 10^6} = 0,1755 \rightarrow 0,38 \lambda_2 \cong 6,669 \text{cm} \quad (3.7)$$

Verifica-se que o valor de 15 cm é suficiente para garantir que os sinais em cada ponto da malha quadrangular estão descorrelacionados.

A tabela 1 que sintetiza os principais parâmetros do sistema de sondagem.

3.2.4 Coleta oficial de dados

Uma vez que os equipamentos estavam devidamente inspecionados e calibrados, as aquisições dos sinais transmitidos durante as fases de medições oficiais passaram a refletir somente o comportamento dos canais de propagação por onde os sinais propagaram.

Posto que a técnica utilizada na sondagem do canal rádio banda larga foi a varredura em frequência (*swept frequency – chirp*), para cada varredura em frequência de 696ms, obteve-se, no domínio dos retardos, um conjunto de respostas impulsivas com 1601 amostras separadas de τ , sendo a primeira amostra em visada direta (τ igual à zero) e a última amostra correspondendo ao maior valor de retardo que se conseguiu medir (τ igual à $\tau_{\text{máx}}$).

Para cada ambiente, os equipamentos foram montados e calibrados para colher os sinais em diferentes localidades de recepção, nomeadas de RX. Como exemplo, veja abaixo a figura 24 do ambiente de medição A5, hall do prédio novo de engenharia da UFF com sete localidades de medição (de RX1 a RX7), no campus da Praia Vermelha. Para cada uma das sete localidades, foi montada uma malha quadrangular seis por seis (6x6),

com trinta e seis (36) posições de recepção espaçadas igualmente de quinze centímetros (15 cm) [38]. O uso de tal malha na recepção se propôs a obtenção de medições independentes em uma mesma localidade RX. Para tal, a mesma antena receptora foi posicionada sequencialmente em cada ponto da malha quadrangular, todas as cautelas descritas acima foram adotadas para garantir a estacionariedade do canal, durante cada varredura de sinal e em cada ponto da malha quadrangular. As medições oficiais foram realizadas sucessivamente, de ponto em ponto até completar todos os trinta e seis pontos da malha quadrangular de cada localidade RX.



Figura 24- Ambiente de medição A5 [38].

Ao final da campanha de medições, obtiveram-se os dados de medição de diferentes localidades em diferentes ambientes. As medições oficiais seguiram o mesmo padrão para a grande maioria dos ambientes, englobando os ambientes A5, A6, A7, A8, B3, B4, C1 e C4 adotados no presente estudo. Ou seja, para cada localidade de cada ambiente, montaram-se malhas quadrangulares seis por seis (6x6) com quinze centímetros (15 cm) de espaçamento. Mantiveram-se os padrões para garantir a estacionariedade e coletaram-se os dados para posterior análise dos canais nos diferentes ambientes de medições.

Capítulo 4 - Metodologia proposta

O método estatístico HMM possui um importante papel na modelagem de sistemas dinâmicos [35]. Ele apresenta grande robustez, com algoritmos iterativos que acabam por ajustar automaticamente os parâmetros do modelo. Por isso, tem sido amplamente estudado e adotado em diferentes modelagens. O HMM tem satisfatória aplicação em sinais que variam com o tempo, pois assegura, em sua natureza probabilística e adaptativa, a maior vantagem para caracterizar sinais corrompidos por ruídos, tais como a fala, a escrita e o próprio sinal rádio.

Há diferentes métodos disponíveis para a modelagem do CRM. Nesse trabalho, usaremos o HMM. Para tanto, modelos e metodologias precisaram ser adotados, os dados foram coletados e pré-processados para o posterior uso do HMM.

4.1 Adoção do canal como QWSSUS

Embora se observe essa resposta aleatória do CRM com variação temporal, pode-se incorporar uma natureza estacionária no tempo ao acatar o processo de caracterização do CRM como um encadeamento de curtos espaços temporais, transformando-se em um somatório de pequenos intervalos com características estatísticas comuns, que apresentam estabilidade, evolução e estacionariedade temporal.

Compreendendo esses pequenos trechos temporais como estacionários, pode-se considerar o sistema como um processo estocástico discreto (as v.a. assumem valores discretos) e aproximadamente estacionário no sentido amplo. Ou seja, existe a delimitação de curtos intervalos temporais onde as propriedades estatísticas do canal apresentam características ditas quase estacionárias no sentido amplo. Melhor dizendo, para curtos intervalos temporais entre 2 instantes, supõe-se o valor esperado como constante e a autocorrelação como função apenas da diferença entre esses 2 instantes.

O canal foi considerado aproximadamente estacionário no sentido amplo (QWSSUS). Dessa forma, considera-se que o canal de propagação pode ser dividido em pequenos setores temporais no interior dos quais se considera que o nível médio da intensidade do sinal propagante é constante e as características estatísticas são comuns. Esses setores temporais são fixos e correspondem a uma “varredura” de sinal no canal,

dada à técnica de varredura em frequência empregada. Assim, a função de transferência de tempo variante do canal $T(f, t)$ é composta apenas de amostras discretas de amplitude referentes as respostas em frequência dentro da banda sondada.

4.2 Aquisição e pré-processamento dos dados

Os sinais recebidos, medidas consecutivas de amplitudes e fases da resposta em frequência do canal, correspondem às respostas do canal frente à varredura das portadoras senoidais transmitidas e, dessa forma, dizem respeito à função de transferência do canal $T(f, t)$ definida como a resposta em frequência do CRM para cada curto intervalo temporal [24].

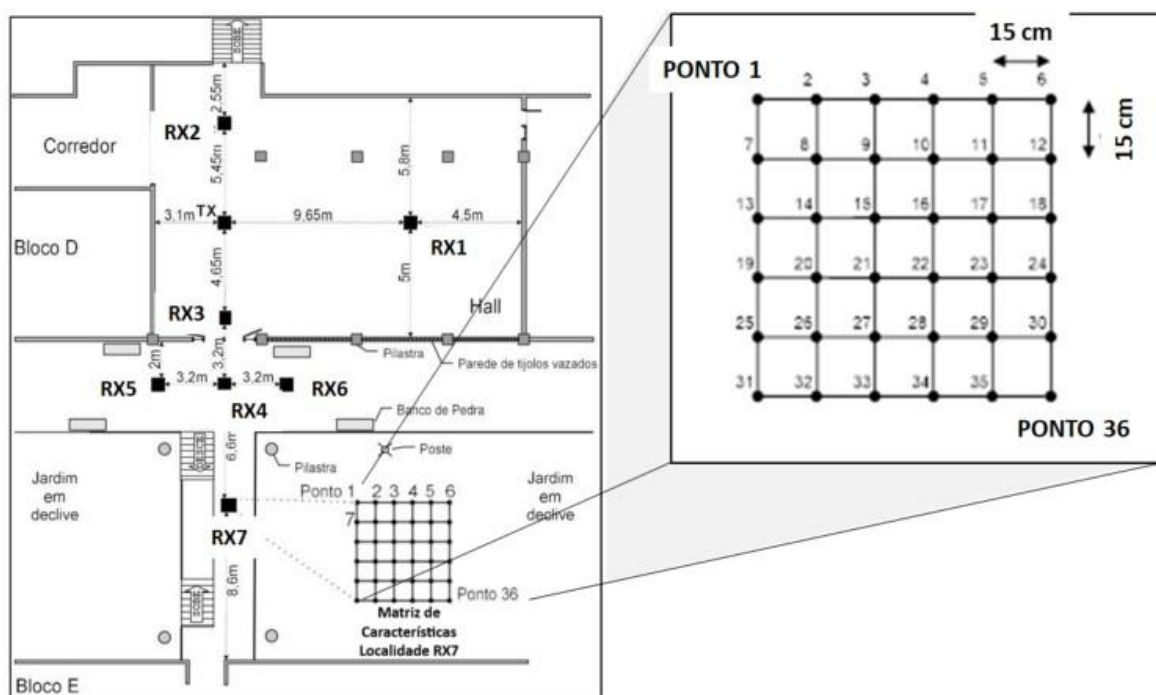


Figura 25- Planta detalhada do ambiente de medição A5.

Para o ambiente de medição A5, a recepção foi montada para coletar $T(f, t)$ em sete diferentes localidades de medição, variadas de RX1 a RX7, mostradas na figura 25. Para cada localidade foi montada uma malha quadrangular 6 por 6, com 36 posições de recepção espaçadas igualmente de 15 cm [38]. Como $0,38 \lambda$ representa o primeiro zero de correlação entre medições [24], a distância de 15 cm foi suficiente para garantir que as

medições das 36 posições de cada localidade não estivessem correlacionadas. Sendo a banda de frequência larga, os valores de λ variaram entre 17,54 cm (frequência superior de 1710 MHz) e 31,25 cm (frequência inferior de 960 MHz) e $0,38 \lambda$ corresponde, respectivamente, a 6,7 cm e 11,9 cm. De onde se conclui que o espaçamento de 15 cm entre os pontos da malha foi suficiente para assegurar a descorrelação entre os sinais medidos nos 36 pontos. As malhas permitiram o levantamento de estatísticas de pequena escala do CRM em cada uma das 36 posições de cada uma das sete localidades de medição.

Ao longo de cada varredura, de 696 ms para o equipamento empregado, realizada em cada posição de recepção fixa, praticamente não havia variação temporal do sinal recebido e a função de transferência de tempo variante $T(f, t)$ obtida estava composta apenas de amostras discretas de amplitudes e fases, conforme mostrado na figura 26.

	1	2	3	4	5	6	7	8	9	10
Canal 1	-26.8722	-26.4155	-25.9509	-25.6850	-25.3852	-25.2302	-24.6744	-24.3059	-23.9446	-23.7799
Canal 2	-129.2281	-136.2621	-145.5318	-153.8580	-158.5876	-165.2975	-169.5081	-174.9092	177.8645	171.5666

Figura 26- Primeiro ponto de RX1 para o ambiente de medição A5.

O analisador vetorial, utilizado como transeptor, colheu 1601 amostras discretas de amplitudes e fases. Assim, o canal 1 registrou as respostas de amplitude e o canal 2 registrou as respostas de fase. A variação em fase do sinal segue uma distribuição uniforme $[-\pi, \pi]$.

```

a0 = 0.42323; % Coeficientes 1 Janela Blackman Harris c/ 3 termos
a1 = 0.49755; % Coeficientes 2 Janela Blackman Harris c/ 3 termos
a2 = 0.07922; % Coeficientes 3 Janela Blackman Harris c/ 3 termos
% Vetor = blackmanharris(pontos, 'periodic'); % Array/matriz coluna
% wvtool(Vetor)
% PBH = Vetor'; % Gera array/matriz linha a partir do array Vetor
%
% Função blackmanharris() % Minimum 4-term Blackman-Harris window
% Janela Blackman Harris c/ 4 termos (a0, a1, a2 e a3)
% a0 = 0.35875, a1 = 0.48829, a2 = 0.14128, a3 = 0.01168
%
for n = 0:(pontos-1)
    Vetor(n+1) = a0-a1*cos(2*pi*n/pontos)+a2*cos(2*pi*2*n/pontos);
end
PBH = Vetor'; % Gera array com 1601 medições armazenadas em Vetor

```

3 termos Blackman-Harris

Função Blackman-Harris

Figura 27- Função janela A5 Blackman-Harris aplicada no Matlab.

Sendo o canal considerado como um filtro linear, os registros de entrada e saída caracterizam diretamente a função de transferência de tempo variante $T(f, t)$. Usou-se a

mesma função janela Blackman-Harris de 3 termos, para minimizar o vazamento espectral provocado pela discretização, de acordo com o estudo realizado por Souza [38]. Tal janela também foi selecionada por Ribeiro [21]. A figura 27 mostra a parte do programa onde se define essa janela.

Aplicou-se a Transformada Discreta Inversa de Fourier (IDFT) para obtenção dos perfis de potência de retardo $P_h(\tau)$ [24]. Nesses perfis, foi aplicada a técnica CFAR (*Constant False Alarm Rate*) para limpeza de ruído [39], considerado taxa de multipercursos válidos de 5% [21]. Após essa aplicação, apenas os percursos válidos foram evidenciados, através de suas amplitudes e retardos, conforme mostrado na figura 28.

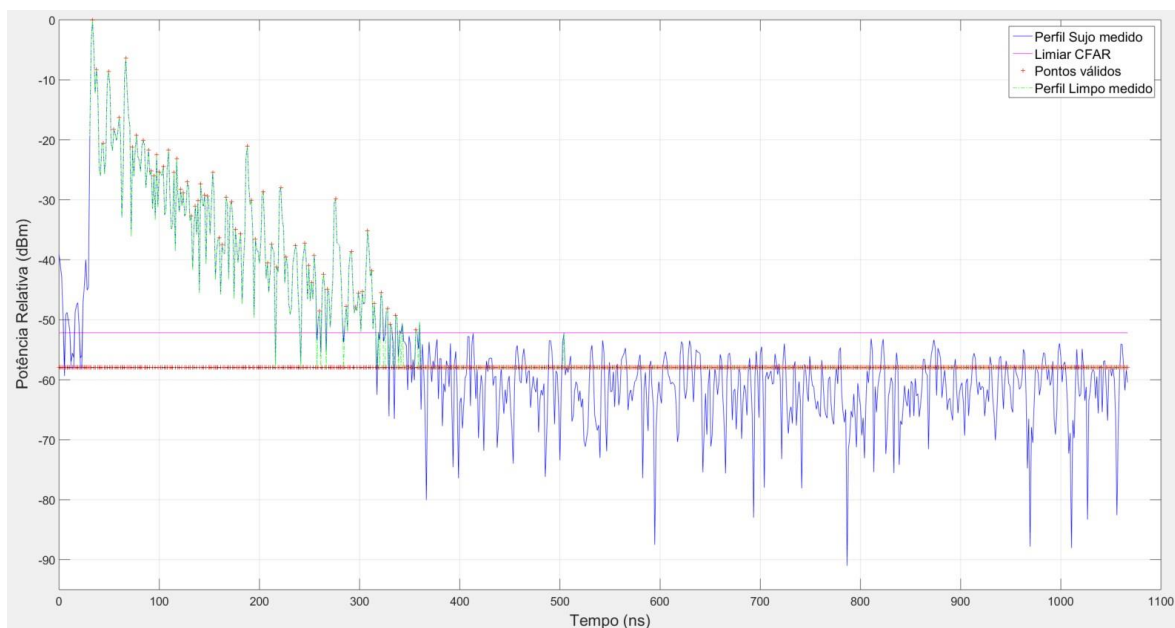


Figura 28- Aplicação da técnica CFAR para limpeza de ruído.

Após o pré-processamento desses dados, obtiveram-se os parâmetros para caracterização do CRM: retardo médio (τ_m) e espalhamento de retardo (σ_T) [24]. Dessa forma, para cada posição da malha quadrangular de cada localidade de medição (RX1 a RX7), geraram-se matrizes com valores de τ_m e σ_T . Essas formatações das amostras em matrizes bidimensionais do tipo 36 x 2 foram denominadas de matrizes de características por localidade e sua representação pode ser vista na figura 29.

	1	2
1	36.9931	24.1921
2	43.5559	22.5312
3	19.6919	24.3349
	⋮	
34	38.7957	22.6824
35	39.5240	22.8705
36	37.3699	21.3394

Figura 29- Representação da matriz de características por localidade de RX1.

4.3 Adoção da topologia de Bakis

Parte importante da modelagem adotada diz respeito à forma como o HMM percorreria os diferentes estados e perguntas importantes, inerentes ao desenvolvimento do estudo em questão, necessitam de respostas. Questionamentos, tais como:

- De que maneira ocorrerão as transições entre estados?
- Essas transições serão determinísticas e pré-estabelecidas ou seguirão de forma arbitrária entre os estados?
- As transições serão completamente aleatórias entre todos os estados ou haverá critérios de transições?
- Será adequado pensar em uma estrutura temporal para a topologia adotada?
- Serão adotadas eficiências em termos de processamentos dos dados relativos ao HMM?

Em vista das topologias detalhadas na seção 2.9, buscou-se a adoção de uma que apresentasse eficiência para as iterações e treinamentos necessários ao HMM e que ainda mantivesse certa flexibilidade. Visto que a execução prática impõe limites ao tamanho do modelo [16] [35], considerou-se, na medida do possível, a devida restrição no número de caminhos a serem percorridos, ou seja, buscou-se sempre um equilíbrio entre flexibilidade e eficiência para a configuração adotada.

Logo, em vista do exposto acima, adotou-se a topologia de Bakis [15] [16]. A escolha foi motivada pelo fato de ser uma simplificação da topologia esquerda-direita, por apresentar uma estrutura de progressão gradual com o tempo que se encaixa bem ao

modelo em questão e por demonstrar adequada eficiência e flexibilidade para as iterações e treinamentos necessários.

Conforme ressaltado na seção 2.9.3, a topologia de Bakis permite a permanência no estado atual, a transição para o estado seguinte ou o salto de apenas um estado subsequente. Com esta estrutura temporal, foi possível migrar entre os sete estados previamente listados (de RX1 até RX7), mantendo certa flexibilidade e eficiência, como se vê na figura 30.

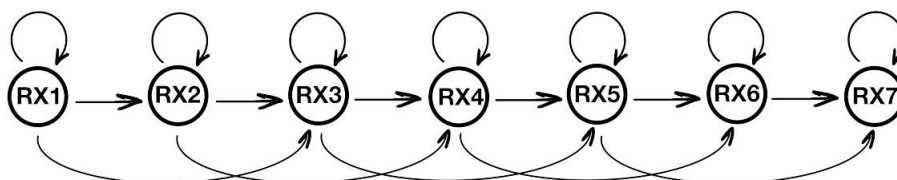


Figura 30- Adoção da topologia de Bakis para transição entre estados.

Para tanto, a matriz de probabilidade de transição de estados A exibiu valores não nulos apenas para a_{ij} onde $j = i$, $j = i + 1$ ou $j = i + 2$. Dessa forma, a matriz A_{Bakis} apresentou uma estrutura triangular superior, visto que qualquer um dos sete estados transicionou apenas para possíveis estados posteriores e a matriz ficou definida conforme a matriz abaixo.

$$A_{Bakis} = \begin{bmatrix} a_{11} & a_{12} & a_{13} & 0 & 0 & 0 & 0 \\ 0 & a_{22} & a_{23} & a_{24} & 0 & 0 & 0 \\ 0 & 0 & a_{33} & a_{34} & a_{35} & 0 & 0 \\ 0 & 0 & 0 & a_{44} & a_{45} & a_{46} & 0 \\ 0 & 0 & 0 & 0 & a_{55} & a_{56} & a_{57} \\ 0 & 0 & 0 & 0 & 0 & a_{66} & a_{67} \\ 0 & 0 & 0 & 0 & 0 & 0 & a_{77} \end{bmatrix} \quad (4.1)$$

4.4 Caracterização do CRM com uso do HMM

Após a aquisição dos parâmetros de retardo médio (τ_m) e espalhamento de retardo (σ_T) para cada localidade e respectivo armazenamento nas matrizes de características por localidade, era necessário adequar os dados de sondagem do CRM ao funcionamento do *toolkit* do HMM para Matlab [40].


```

Rx1ToRx2data
2x2x36 double
val(:, :, 1) =
  9.5790    25.2319
 31.4779    28.1383
val(:, :, 2) =
  8.8331    22.7515
 29.4265    24.5623
      ⋮
val(:, :, 35) =
 40.4554    25.2335
 24.5487    25.1018
val(:, :, 36) =
 41.1742    24.7449
  6.6438    23.4707

```

Figura 31- Representação da matriz de características de transição.

Para tal adequação, as matrizes de características das localidades RX1 a RX7 correspondem aos estados. As transições entre duas localidades, conforme a topologia Bakis, permitiram agrupar os parâmetros de dispersão τ_m e σ_T do CRM em matrizes tridimensionais do tipo $2 \times 2 \times 36$, aleatoriamente reunindo dois a dois os 36 parâmetros τ_m e σ_T de cada localidade, como visto na figura 31. Essas matrizes tridimensionais foram denominadas de matrizes de características de transição e, no contexto do HMM, correspondem aos observáveis. A figura 32 representa o uso do HMM para caracterização do CRM, mostrando, respectivamente, as matrizes de características por localidade e as matrizes de características de transição como estados e observáveis do HMM.

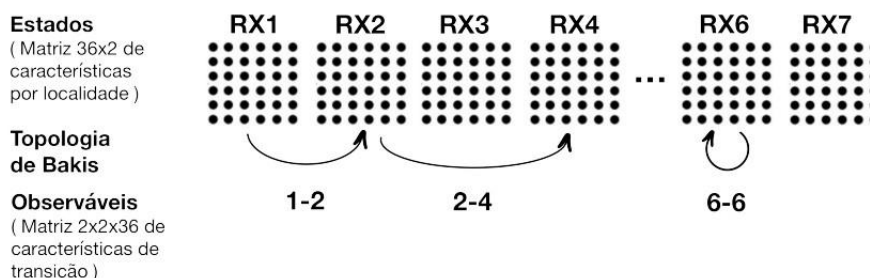


Figura 32- Adoção do HMM e da topologia de Bakis para caracterização do CRM.

Quando as matrizes de características de transição foram agrupadas, indicando a passagem de um estado para outro, indo de RX1 até RX7, segundo a topologia Bakis,

formou-se a matriz base de características do CRM com quatro dimensões. A figura 33 representa uma matriz base de características gerada a partir de 9 transições.

MATRIZ BASE DE CARACTERÍSTICAS					
4-D double					
val (:, :, 1, 1) =		val (:, :, 1, 2) =		val (:, :, 1, 9) =	
9.5790	25.2319	29.0055	23.4865	47.3199	31.8564
31.4779	28.1383	29.6433	25.3508	41.2293	25.7735
	⋮		⋮		⋮
val (:, :, 36, 1) =		val (:, :, 36, 2) =		val (:, :, 36, 9) =	
41.1742	24.7449	6.0538	22.9225	35.0224	23.2184
6.6438	23.4707	29.7063	27.7049	45.6587	25.7165

Figura 33- Representação da matriz base de características.

Essa matriz base foi dividida ao meio entre matriz base de treinamento e matriz base de teste. Dessa forma, 50% dos dados foram utilizados para inicialização e treinamento com o algoritmo Baum-Welch [21], enquanto outros 50% foram tratados na seguinte sequência: obtiveram-se as medidas de log-verossimilhança para todas as amostras da matriz base de teste, retiraram-se os valores discrepantes em relação aos demais (*outliers*), separaram-se igualmente os dados em grupo de referência (25% para o cálculo dos limiares criados) e grupo de teste (25% para testar os limiares). A figura 34 representa graficamente os passos descritos acima para aplicação do HMM ao CRM usando o *toolkit* do HMM para Matlab.

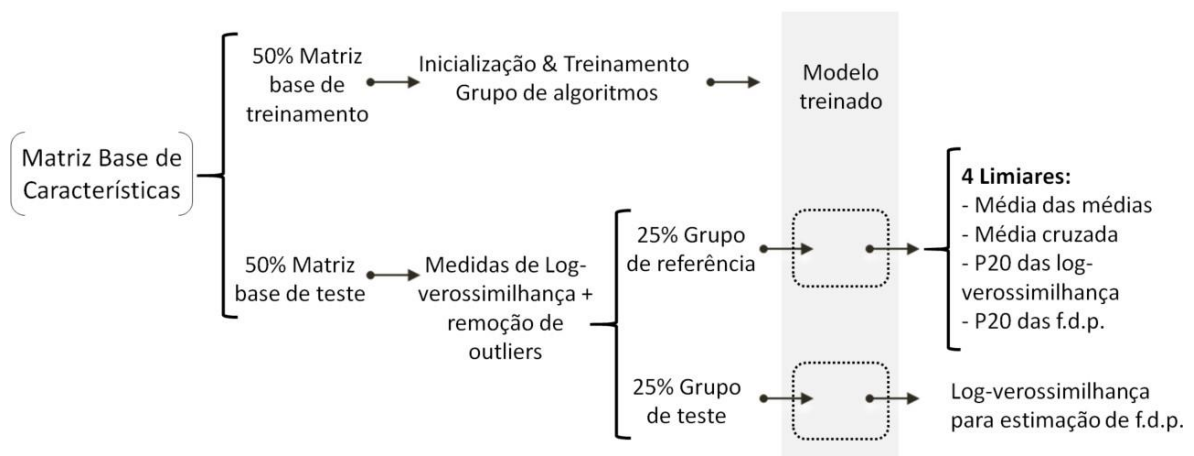


Figura 34- Representação dos passos para aplicação do HMM ao CRM.

Para treinar o HMM na caracterização do CRM, consideraram-se as etapas de inicialização e treinamento. Para inicialização, efetuaram-se varreduras para descobertas de valores ótimos. Ainda no processo de inicialização, supondo que cada estado costuma gerar grupos de observáveis tipicamente próximos, executou-se análise de agrupamento (*clustering*) com o algoritmo iterativo *k-means* para reunir vetores de observáveis semelhantes entre si e para associar esses vetores às misturas gaussianas dos estados.

Cabe, aqui, uma apropriada descrição sobre o agrupamento e o algoritmo *k-means*. A busca inicial por padrões começa na busca por similaridades entre os dados de forma a agrupá-los em classes ou subconjuntos. Sendo o algoritmo EM (de Baum-Welch) sensível em sua fase de inicialização dos parâmetros, usou-se o algoritmo *k-means* (baixo custo computacional) para análise dos dados e definição adequada do número de agrupamentos (*clusters*). Usaram-se os centros finais como as médias iniciais do algoritmo EM.

O algoritmo *k-means* associa cada observável ao agrupamento que tem média com a menor distância euclidiana e, a partir dessa associação, recalcula as médias dos agrupamentos. Em cada etapa são recalculadas as misturas gaussianas e todos os parâmetros do modelo M_{HMM} e a iteração é feita até que a distância entre o novo HMM e o antigo seja menor que o limite arbitrado.

Ainda na etapa de treinamento, usou-se o algoritmo Baum-Welch para treinamento e maximização dos parâmetros do modelo. Com base nos valores iniciais, fez-se o cálculo dos caminhos mais prováveis (*expectation*) e o recálculo das probabilidades com base nos novos caminhos calculados anteriormente (*maximization*). Esse processo gerou novos valores iniciais que foram utilizados no lugar dos antigos. Dessa forma, criaram-se iterações até que os cálculos alcançassem um critério de parada, que poderia ser pouca melhoria na verossimilhança em comparação com os parâmetros anteriores, ou número limite de iterações, ou, ainda, uma combinação entre os dois critérios anteriores.

Para a matriz base de teste, com relação ao grupo de referência, os dados foram usados para a definição de limiares. Esses correspondem aos valores de log-verossimilhanças usados como referências, ou seja, para cada matriz de características de transição de um estado para outro, serão log-verossimilhanças válidas os valores do grupo de teste que estão acima do limiar analisado. Foram definidos quatro diferentes limiares, descritos na tabela 2, com os quais os grupos de teste foram comparados [21].

Para a matriz base de teste, com relação ao grupo de teste, os valores de log-verossimilhanças foram usados para estimação de funções de densidade de probabilidade

(comando *ksdensity*). Para cada transição entre estados, obteve-se uma função densidade de probabilidade (f.d.p.).

Tabela 2 - Especificação dos limiares escolhidos [21].

Limiares	Especificações
1. Média das médias	Média aritmética simples entre a média das log-verossimilhanças da localidade em questão e demais localidades.
2. Média Cruzada	Média das log-verossimilhanças da localidade em questão é multiplicada pelo desvio-padrão das log-verossimilhanças das demais localidades. Média das log-verossimilhanças das demais localidades é multiplicada pelo desvio-padrão das log-verossimilhanças da localidade em questão. Calcula-se a média aritmética simples dos dois valores obtidos.
3. Percentil P20 das log-verossimilhanças	Limiar é posicionado de modo a ser maior que 20% das log-verossimilhanças da localidade em questão e demais localidades.
4. Percentil P20 das f.d.p.	Limiar separa 20% da área sob a f.d.p. das log-verossimilhanças à sua esquerda e os demais 80% à sua direita.

Em posse dos valores dos limiares e das log-verossimilhanças, as simulações foram feitas e os resultados foram obtidos através de:

- Análise gráfica de f.d.p. *versus* limiares calculados. Cada gráfico indica os 4 limiares calculados e as f.d.p., com base no HMM treinado para cada transição de estados.
- Análise dos modelos HMM treinados. Usou-se uma matriz de confusão para classificar o posicionamento das log-verossimilhanças em relação aos limiares calculados para cada transição de estados.

Capítulo 5 - Simulações e resultados

5.1 Simulação 1: transições aleatórias entre os estados do mesmo ambiente de medição A5 (*indoor-outdoor*)

Para a simulação 1, a topologia de Bakis gerou transições aleatórias de estados, ou seja, o programa em Matlab gerou aleatoriamente as transições entre as localidades RX1 e RX7. Como exemplo, foram geradas 7 transições:

- RX1 para RX1 (transição 1 representado por 1-1).
- RX1 para RX3 (transição 2 representado por 1-3).
- RX3 para RX4 (transição 3 representado por 3-4).
- RX4 para RX5 (transição 4 representado por 4-5).
- RX5 para RX5 (transição 5 representado por 5-5).
- RX5 para RX7 (transição 6 representado por 5-7).
- RX7 para RX7 (transição 7 representado por 7-7).

Cada transição entre localidades definiu uma matriz de características de transição que permitiu agrupar os parâmetros de dispersão temporal τ_m e σ_T das respectivas localidades em matrizes tridimensionais do tipo $2 \times 2 \times 36$. Metade das amostras de cada matriz de características de transição foi usada para treinamento do HMM e a outra metade foi destinada para teste e validação. Para inicialização, atribuiu-se o valor 2 ao número de estados, pois a análise foi feita para cada transição. Para as misturas gaussianas e o número de iterações, efetuaram-se varreduras para descobrir os valores ótimos e esses foram, respectivamente, 2 e 20.

Com a metade reservada para teste e validação, seguiram-se os passos:

- Valores de log-verossimilhanças foram calculados com base em todo o conjunto de dados (todo o conjunto de teste e validação).
- 2 valores de *outliers* foram retirados para cada matriz de características de transição.
- 50% das amostras de teste e validação foram usadas para calcular limiares.
- 50% das amostras de teste e validação foram usadas para testar os limiares calculados.

O exame dos resultados, com base na análise gráfica, pode ser visualizado na figura 35. A linha vermelha corresponde aos valores de log-verossimilhanças da transição 3-4, escolhida como exemplo para ilustração. As demais curvas de densidade de probabilidade correspondem aos valores de log-verossimilhanças das outras 6 transições. Percebe-se que o limiar 3, que corresponde ao percentil P20 das log-verossimilhanças, é o que apresenta melhor ajuste na identificação das log-verossimilhanças da transição 3-4. Isso porque os valores de log-verossimilhanças da transição 3-4, praticamente, encontram-se à direita do limiar 3 e os demais valores encontram-se à esquerda do mesmo.

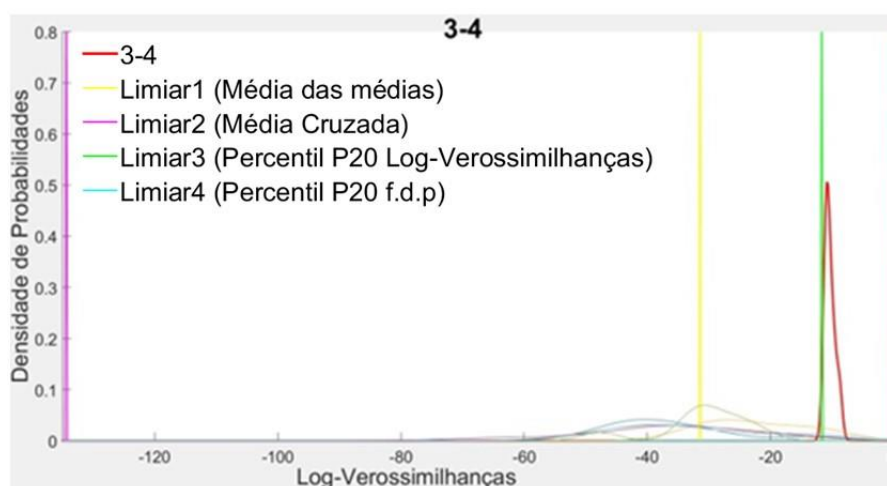


Figura 35- Estimação de f.d.p. para a transição 3-4 do ambiente A5.

Comportamento análogo foi observado para os gráficos das demais 6 transições (1-1, 1-3, 4-5, 5-5, 5-7, 7-7). De onde se inferiu que o limiar 3 (percentil P20 das log-verossimilhanças) é a melhor escolha entre as quatro opções consideradas.

Tabela 3 - Matriz de confusão para cada transição de estados de A5.

Transição de Estados	Verdadeiro Positivo	Falso Negativo	Verdadeiro Negativo	Falso Positivo
1-1	12,5% (7/8)	1,8% (1/8)	69,6% (39/48)	16,1% (9/48)
1-3	10,7% (6/8)	3,6% (2/8)	82,1% (46/48)	3,6% (2/48)
3-4	14,3% (8/8)	0% (0/8)	83,9% (47/48)	1,8% (1/48)
4-5	10,7% (6/8)	3,6% (2/8)	46,4% (26/48)	39,3% (22/48)
5-5	8,9% (5/8)	5,4% (3/8)	1,8% (1/48)	83,9% (47/48)
5-7	10,7% (6/8)	3,6% (2/8)	35,7% (20/48)	50,0% (28/48)
7-7	12,5% (7/8)	1,8% (1/8)	78,6% (44/48)	7,1% (4/48)

O mesmo resultado foi apresentado quando se analisou a matriz de confusão gerada. Levando-se em conta apenas o limiar 3, limiar com melhor resultado, foi possível montar uma matriz de confusão para cada transição de estados. Essa matriz de confusão pode ser vista na tabela 3.

No total, foram testados 56 valores de log-verossimilhança. Oito deles foram valores de teste de log-verossimilhança da matriz de características de transição que se estava analisando, e 48 foram valores das demais 6 transições de estados. Observa-se, na tabela 3, que as porcentagens das colunas verdadeiro positivo e falso negativo dizem respeito apenas aos valores da transição de estados corrente. Pegando como exemplo a própria transição de estados 3-4, o HMM acertou todos os 8 valores testados (não apresentou falsos negativos).

O mesmo raciocínio é visto para as colunas verdadeiro negativo e falso positivo. Quando se analisa uma determinada transição, essas colunas correspondem aos valores das outras 6 transições em relação a transição analisada. Assim, para a transição de estados 3-4, é possível observar que o HMM rejeitou acertadamente 47 de 48 valores de log-verossimilhança.

5.2 Simulação 2: transições determinadas entre os estados do mesmo ambiente de medição A5 (*indoor-outdoor*)

Ao passo que a simulação 1 analisou cada transição individualmente, a simulação 2 propôs-se a analisar o conjunto de transições como um todo. Para isso, a topologia de Bakis gerou transições pré-definidas de estados, ou seja, definiu-se, de antemão, como ocorreriam as transições entre as localidades RX1 e RX7. Foram estabelecidas 9 transições:

- RX1 para RX2 (transição 1 representado por 1-2).
- RX2 para RX2 (transição 2 representado por 2-2).
- RX2 para RX3 (transição 3 representado por 2-3).
- RX3 para RX4 (transição 4 representado por 3-4).
- RX4 para RX4 (transição 5 representado por 4-4).
- RX4 para RX5 (transição 6 representado por 4-5).

- RX5 para RX6 (transição 7 representado por 5-6).
- RX6 para RX6 (transição 8 representado por 6-6).
- RX6 para RX7 (transição 9 representado por 6-7).

Uma matriz de características de transição do tipo $2 \times 2 \times 36$ foi gerada para cada uma das 9 transições acima mencionadas, agrupando os parâmetros de dispersão temporal τ_m e σ_T das respectivas localidades. Essas matrizes foram reunidas, formando uma matriz base de características do tipo $2 \times 2 \times 36 \times 9$. Metade das amostras da matriz base de características foi reservada para treinamento do HMM e a outra metade foi destinada para teste e validação. Cabe ressaltar que essa divisão foi feita para cada transição, de modo que tanto a base de treinamento quanto a de teste e validação agrupassem dados de todas as transições. Para inicialização, atribuiu-se valor 10 ao número de estados. Para as misturas gaussianas e número de iterações, fez-se novas varreduras com valores ótimos 2 e 20, respectivamente. Isto posto, para a metade atribuída ao conjunto de teste e validação, seguiu-se com o cálculo dos valores de log-verossimilhança, retirada de *outliers*, divisão das amostras para determinação e teste dos limiares.

O exame dos resultados, com base na análise gráfica, pode ser visto na figura 36. A linha vermelha corresponde aos valores de log-verossimilhança do conjunto total de transições. Percebe-se que o limiar 3 ainda é o que apresenta melhor ajuste na identificação das log-verossimilhanças válidas, uma vez que esses valores de log-verossimilhança, em sua maioria, encontram-se à direita do limiar 3.

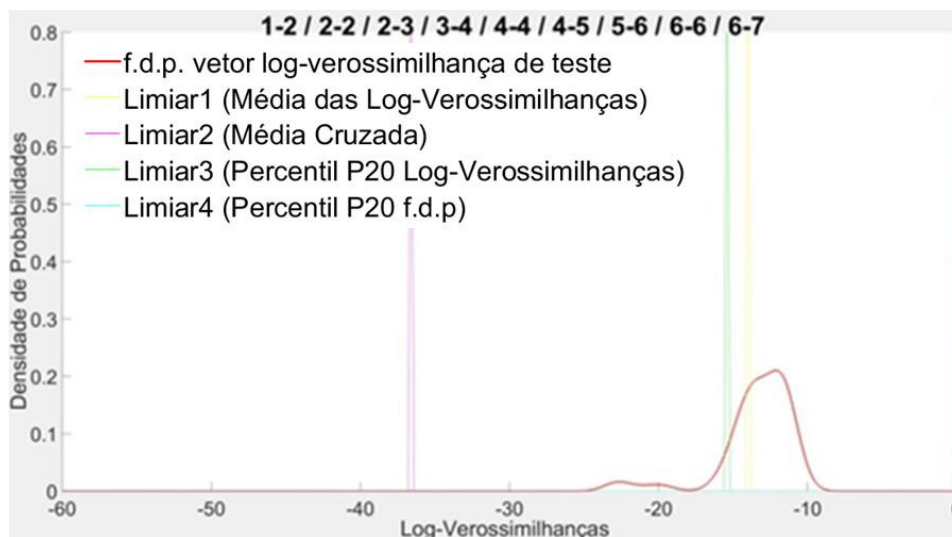


Figura 36- Estimação de f.d.p. para 9 transições do ambiente A5.

Nota-se que não existiram as demais curvas de densidade de probabilidade, visto que a análise aqui é feita com toda a matriz base de características (conjunto total com 9 transições). Além disso, para a matriz de confusão, não existiram dados para as colunas verdadeiro negativo e falso positivo, visto que não houve comparações entre diferentes transições.

Como feito na simulação 1, levando-se em conta apenas o limiar 3, foi possível montar a matriz de confusão vista na tabela 4. Ao todo, foram testados 79 valores de log-verossimilhança e o HMM acertou 72. Assim sendo, ele rejeitou erroneamente apenas 7 valores testados.

Tabela 4 - Matriz de confusão para o conjunto de transições de A5.

Transição de Estados	Verdadeiro Positivo	Falso Negativo	Verdadeiro Negativo	Falso Positivo
1-2 / 2-2 / 2-3 / 3-4 / 4-4 / 4-5 / 5-6 / 6-6 / 6-7	91,1% (72/79)	8,9% (7/79)	0% (0/79)	0% (0/79)

5.3 Simulação 3: transições determinadas entre os estados do mesmo ambiente de medição A6 (*outdoor-indoor*)

Ao passo que a simulação 2 analisou o conjunto de transições como um todo para o ambiente de medição A5 (*indoor-outdoor*), a simulação 3 propôs-se a analisar também um conjunto com 9 transições, mas agora para o ambiente de medição A6 (*outdoor-indoor*). Para isso, a topologia de Bakis gerou transições pré-definidas de estados, ou seja, definiu-se, de antemão, como ocorreriam as transições entre as localidades RX1 e RX7. As transições ocorreram da mesma forma que na simulação 2. Logo, foram estabelecidas 9 transições:

- RX1 para RX2 (transição 1 representado por 1-2).
- RX2 para RX2 (transição 2 representado por 2-2).
- RX2 para RX3 (transição 3 representado por 2-3).
- RX3 para RX4 (transição 4 representado por 3-4).
- RX4 para RX4 (transição 5 representado por 4-4).

- RX4 para RX5 (transição 6 representado por 4-5).
- RX5 para RX6 (transição 7 representado por 5-6).
- RX6 para RX6 (transição 8 representado por 6-6).
- RX6 para RX7 (transição 9 representado por 6-7).

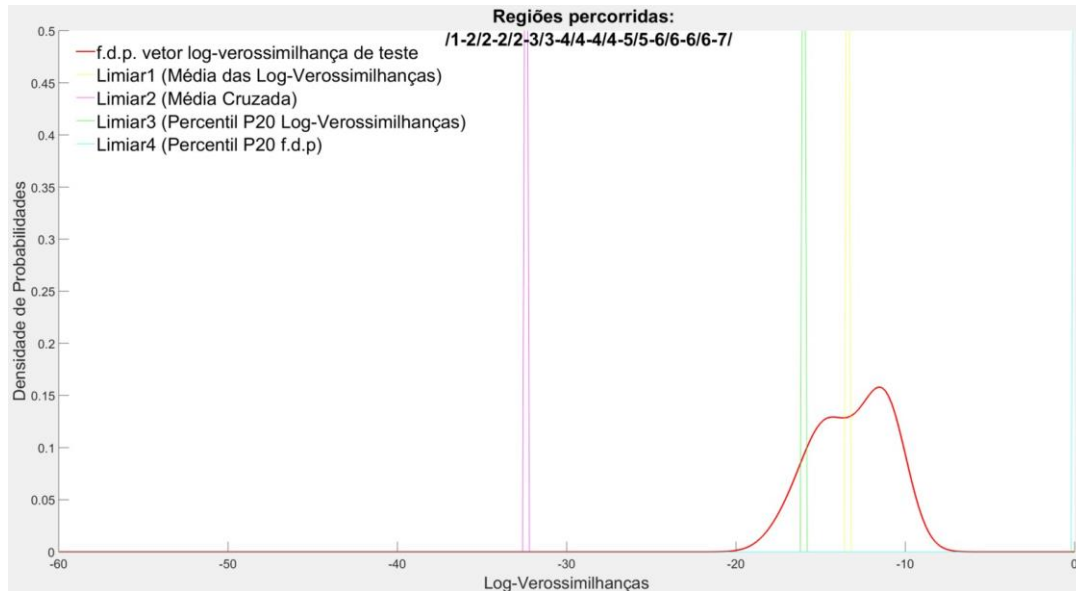


Figura 37- Estimação de f.d.p. para 9 transições do ambiente A6.

Essa simulação 3 seguiu os mesmos passos da simulação 2.

O exame dos resultados, com base na análise gráfica, pode ser visto na figura 37. A linha vermelha corresponde aos valores de log-verossimilhança do conjunto total de transições. Percebe-se que o limiar 3 ainda é o que apresenta melhor ajuste na identificação das log-verossimilhanças válidas, uma vez que esses valores de log-verossimilhança, em sua maioria, encontram-se à direita do limiar 3.

Tabela 5 - Matriz de confusão para o conjunto de transições de A6.

Transição de Estados	Verdadeiro Positivo	Falso Negativo	Verdadeiro Negativo	Falso Positivo
1-2 / 2-2 / 2-3 / 3-4 / 4-4 / 4-5 / 5-6 / 6-6 / 6-7	91,1% (72/79)	8,9% (7/79)	0% (0/79)	0% (0/79)

Nota-se que não existiram as demais curvas de densidade de probabilidade, visto que a análise aqui é feita com toda a matriz base de características (conjunto total com 9 transições). Além disso, para a matriz de confusão, não existiram dados para as colunas

verdadeiro negativo e falso positivo, visto que não houve comparações entre diferentes transições.

Levando-se em conta apenas o limiar 3, foi possível montar a matriz de confusão vista na tabela 5. Ao todo, foram testados 79 valores de log-verossimilhança e o HMM também acertou 72. Assim sendo, ele rejeitou erroneamente apenas 7 valores testados.

5.4 Simulação 4: transições determinadas entre os estados do mesmo ambiente de medição A7 (*indoor-outdoor*)

A simulação 4 propôs-se a analisar um conjunto menor de transições. Posto isso, o ambiente escolhido foi o ambiente de medição A7. A topologia de Bakis gerou transições pré-definidas de estados, ou seja, definiu-se, de antemão, como ocorreriam as transições entre as localidades RX1 e RX4. Foram estabelecidas 3 transições diretas de estados:

- RX1 para RX2 (transição 1 representado por 1-2).
- RX2 para RX3 (transição 2 representado por 2-3).
- RX3 para RX4 (transição 3 representado por 3-4).

Uma matriz de características de transição do tipo $2 \times 2 \times 36$ foi gerada para cada uma das 3 transições acima mencionadas, agrupando os parâmetros de dispersão temporal τ_m e σ_T das respectivas localidades. Essas matrizes foram reunidas, formando uma matriz base de características do tipo $2 \times 2 \times 36 \times 3$. Metade das amostras da matriz base de características foi reservada para treinamento do HMM e a outra metade foi destinada para teste e validação. Cabe ressaltar que essa divisão foi feita para cada transição, de modo que tanto a base de treinamento quanto a de teste e validação agrupassem dados das 3 transições. Para inicialização, atribuiu-se valor 4 ao número de estados. Para as misturas gaussianas e número de iterações, fez-se novas varreduras com valores ótimos 2 e 20, respectivamente. Isto posto, para a metade atribuída ao conjunto de teste e validação, seguiu-se com o cálculo dos valores de log-verossimilhança, retirada de *outliers*, divisão das amostras para determinação e teste dos limiares.

O exame dos resultados, com base na análise gráfica, pode ser visto na figura 38. A linha vermelha corresponde aos valores de log-verossimilhança do conjunto de 3 transições. Percebe-se que o limiar 3 ainda é o que apresenta melhor ajuste na identificação

das log-verossimilhanças válidas, uma vez que esses valores de log-verossimilhança, em sua maioria, encontram-se à direita do limiar 3.

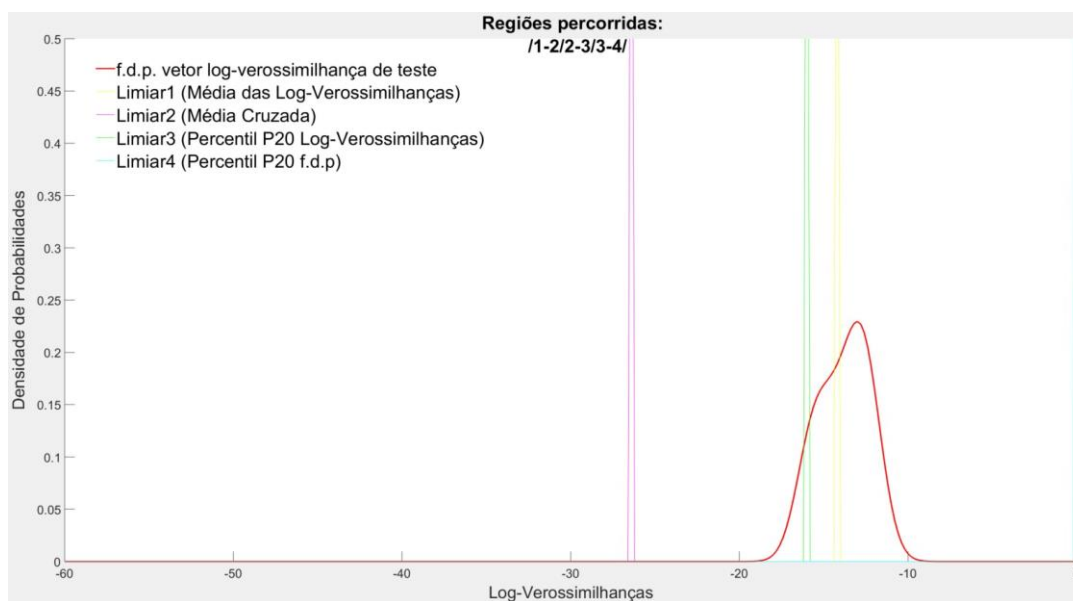


Figura 38- Estimação de f.d.p. para 3 transições do ambiente A7.

Nota-se que não existiram as demais curvas de densidade de probabilidade, visto que a análise aqui é feita com toda a matriz base de características (conjunto total com 3 transições). Além disso, para a matriz de confusão, não existiram dados para as colunas verdadeiro negativo e falso positivo, visto que não houve comparações entre diferentes transições.

Tabela 6 - Matriz de confusão para o conjunto de transições de A7.

Transição de Estados	Verdadeiro Positivo	Falso Negativo	Verdadeiro Negativo	Falso Positivo
1-2 / 2-3 / 3-4	92,0% (23/25)	8,0% (2/25)	0% (0/25)	0% (0/25)

Levando-se em conta apenas o limiar 3, foi possível montar a matriz de confusão vista na tabela 6. Ao todo, foram testados 25 valores de log-verossimilhança e o HMM acertou 23. Assim sendo, ele rejeitou erroneamente apenas 2 valores testados.

5.5 Simulação 5: transições determinadas entre os estados do mesmo ambiente de medição A8 (*outdoor-indoor*)

Ao passo que a simulação 4 analisou um conjunto menor de transições para o ambiente de medição A7 (*indoor-outdoor*), a simulação 5 propôs-se a analisar também um conjunto com 3 transições, mas agora para o ambiente de medição A8 (*outdoor-indoor*). Para isso, a topologia de Bakis gerou transições pré-definidas de estados, ou seja, definiu-se, de antemão, como ocorreriam as transições entre as localidades RX1 e RX4. Foram estabelecidas 3 transições:

- RX1 para RX2 (transição 1 representado por 1-2).
- RX2 para RX3 (transição 2 representado por 2-3).
- RX3 para RX4 (transição 3 representado por 3-4).

Essa simulação 5 seguiu os mesmos passos da simulação 4.

O exame dos resultados, com base na análise gráfica, pode ser visto na figura 39. A linha vermelha corresponde aos valores de log-verossimilhança do conjunto de 3 transições. Percebe-se que os limiares 2 e 3 apresentaram bons ajustes na identificação das log-verossimilhanças válidas, uma vez que esses valores de log-verossimilhança, em sua maioria, encontram-se à direita dos respectivos limiares.

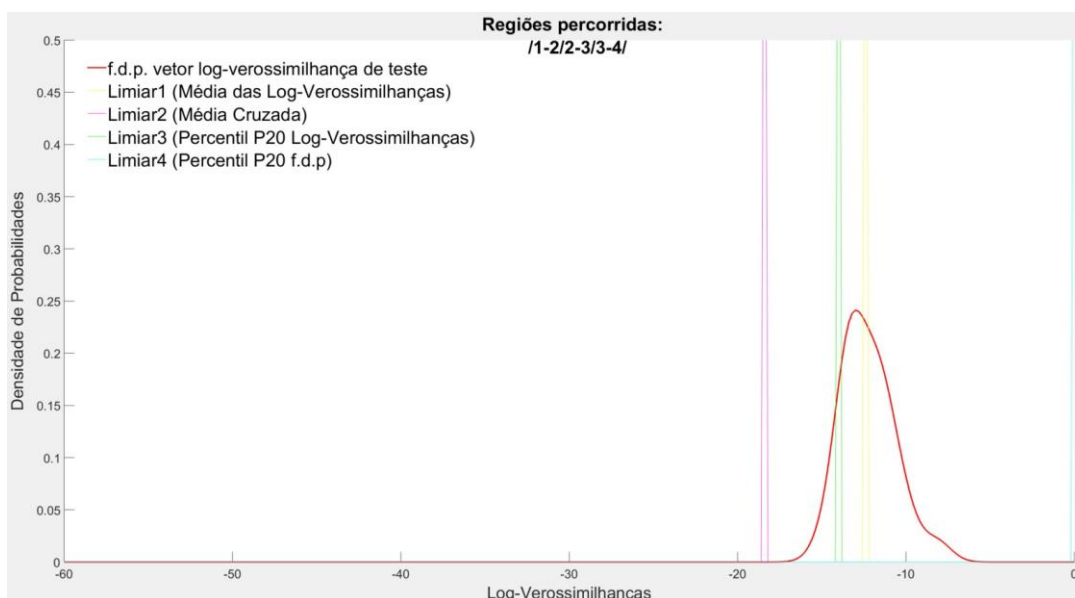


Figura 39- Estimação de f.d.p. para 3 transições do ambiente A8.

De fato, em algumas execuções do programa, o limiar 2 apresentou melhores ajustes que o próprio limiar 3. Dessa forma, o limiar 2 (média cruzada) também será considerado como uma opção válida para a presente simulação.

Nota-se que não existiram as demais curvas de densidade de probabilidade, visto que a análise aqui é feita com toda a matriz base de características (conjunto total com 3 transições). Além disso, para a matriz de confusão, não existiram dados para as colunas verdadeiro negativo e falso positivo, visto que não houve comparações entre diferentes transições.

Tabela 7 - Matrizes de confusão para o conjunto de transições de A8.

Transição de Estados	Verdadeiro Positivo	Falso Negativo	Verdadeiro Negativo	Falso Positivo
1-2 / 2-3 / 3-4	92,0% (23/25)	8,0% (2/25)	0% (0/25)	0% (0/25)

Levando-se em conta apenas o limiar 3, foi possível montar a matriz de confusão vista na tabela 7. Ao todo, foram testados 25 valores de log-verossimilhança e o HMM acertou 23. Assim sendo, ele rejeitou erroneamente apenas 2 valores testados.

Pela análise gráfica, considerando o limiar 2 para a simulação em questão, o HMM acertou todos os 25 valores.

5.6 Simulação 6: transições determinadas entre os estados do mesmo ambiente de medição C4 (outdoor)

A simulação 6 propôs-se a analisar um conjunto também com 3 transições. Os ambientes das simulações anteriores eram ambientes de transição *indoor-outdoor* (A7) e *outdoor-indoor* (A8). Para essa simulação 6, foi considerado o ambiente C4 que é completamente *outdoor*. Posto isso, a topologia de Bakis gerou transições pré-definidas de estados, ou seja, definiu-se, de antemão, como ocorreriam as transições entre as localidades RX1 e RX4. Foram estabelecidas 3 transições diretas de estados:

- RX1 para RX2 (transição 1 representado por 1-2).
- RX2 para RX3 (transição 2 representado por 2-3).
- RX3 para RX4 (transição 3 representado por 3-4).

Essa simulação 6 seguiu os mesmos passos das simulações 4 e 5.

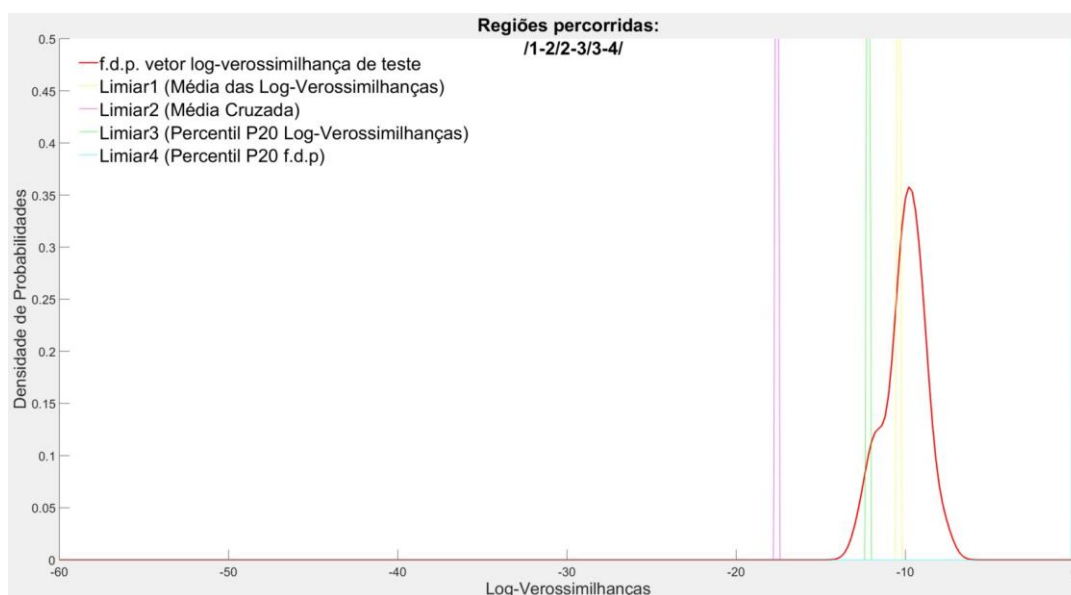


Figura 40- Estimação de f.d.p. para 3 transições do ambiente C4.

O exame dos resultados, com base na análise gráfica, pode ser visto na figura 40. A linha vermelha corresponde aos valores de log-verossimilhança do conjunto de 3 transições. Percebe-se que o limiar 3 ainda é o que apresenta melhor ajuste na identificação das log-verossimilhanças válidas, uma vez que esses valores de log-verossimilhança, em sua maioria, encontram-se à direita do limiar 3.

Nota-se que não existiram as demais curvas de densidade de probabilidade, visto que a análise aqui é feita com toda a matriz base de características (conjunto total com 3 transições). Além disso, para a matriz de confusão, não existiram dados para as colunas verdadeiro negativo e falso positivo, visto que não houve comparações entre diferentes transições.

Tabela 8 - Matrizes de confusão para o conjunto de transições de C4.

Transição de Estados	Verdadeiro Positivo	Falso Negativo	Verdadeiro Negativo	Falso Positivo
1-2 / 2-3 / 3-4	96,0% (24/25)	4,0% (1/25)	0% (0/25)	0% (0/25)

Levando-se em conta apenas o limiar 3, foi possível montar a matriz de confusão vista na tabela 8. Ao todo, foram testados 25 valores de log-verossimilhança e o HMM acertou 24. Assim sendo, ele rejeitou erroneamente apenas 1 valor testado.

5.7 Simulação 7: transições determinadas entre os estados do mesmo ambiente de medição B4 (*outdoor*)

A simulação 7 propôs-se a analisar um conjunto com 4 transições, ou seja, um conjunto um pouco maior que o analisado na simulação anterior. Posto isso, o ambiente de medição escolhido foi o B4 (*outdoor*). Logo, a topologia de Bakis gerou transições pré-definidas de estados, ou seja, definiu-se, de antemão, como ocorreriam as transições entre as localidades RX1 e RX5. Foram estabelecidas 4 transições diretas de estados:

- RX1 para RX2 (transição 1 representado por 1-2).
- RX2 para RX3 (transição 2 representado por 2-3).
- RX3 para RX4 (transição 3 representado por 3-4).
- RX4 para RX5 (transição 4 representado por 4-5).

Uma matriz de características de transição do tipo $2 \times 2 \times 36$ foi gerada para cada uma das 4 transições acima mencionadas, agrupando os parâmetros de dispersão temporal τ_m e σ_T das respectivas localidades. Essas matrizes foram reunidas, formando uma matriz base de características do tipo $2 \times 2 \times 36 \times 4$. Metade das amostras da matriz base de características foi reservada para treinamento do HMM e a outra metade foi destinada para teste e validação. Cabe ressaltar que essa divisão foi feita para cada transição, de modo que tanto a base de treinamento quanto a de teste e validação agrupassem dados das 4 transições. Para inicialização, atribuiu-se valor 5 ao número de estados. Para as misturas gaussianas e número de iterações, fez-se novas varreduras com valores ótimos 2 e 20, respectivamente. Isto posto, para a metade atribuída ao conjunto de teste e validação, seguiu-se com o cálculo dos valores de log-verossimilhança, retirada de *outliers*, divisão das amostras para determinação e teste dos limiares.

O exame dos resultados, com base na análise gráfica, pode ser visto na figura 41. A linha vermelha corresponde aos valores de log-verossimilhança do conjunto de 4 transições. Percebe-se que o limiar 3 apresenta melhor ajuste na identificação das log-

verossimilhanças válidas, uma vez que esses valores de log-verossimilhança, em sua maioria, encontram-se à direita do limiar 3.

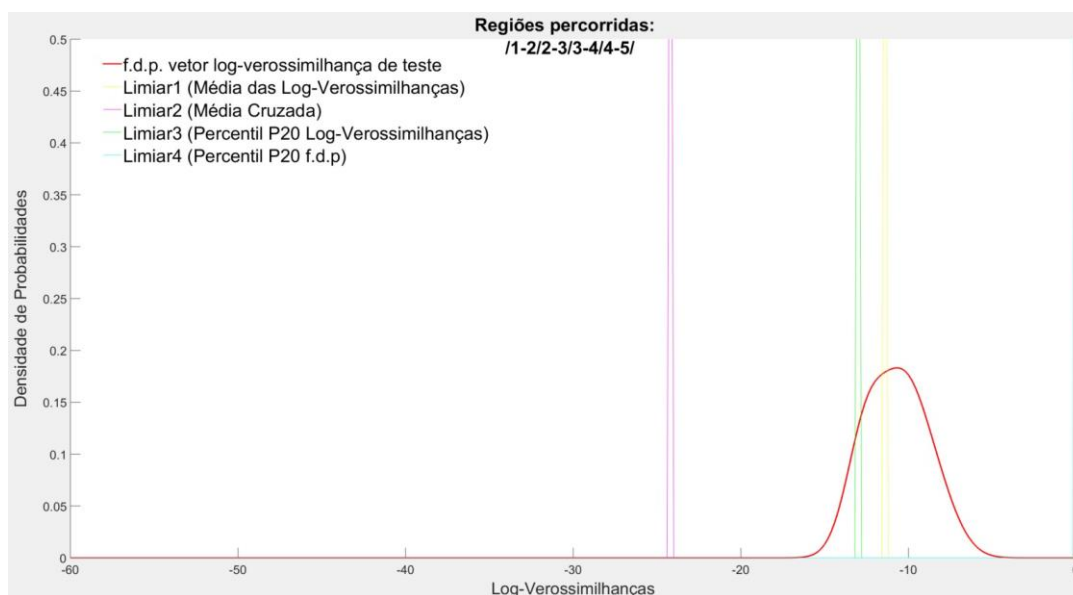


Figura 41- Estimação de f.d.p. para 4 transições do ambiente B4.

Nota-se que não existiram as demais curvas de densidade de probabilidade, visto que a análise aqui é feita com toda a matriz base de características (conjunto total com 4 transições). Além disso, para a matriz de confusão, não existiram dados para as colunas verdadeiro negativo e falso positivo, visto que não houve comparações entre diferentes transições.

Tabela 9 - Matriz de confusão para o conjunto de transições de B4.

Transição de Estados	Verdadeiro Positivo	Falso Negativo	Verdadeiro Negativo	Falso Positivo
1-2 / 2-3 / 3-4/ 4-5	94,12% (32/34)	5,88% (2/34)	0% (0/34)	0% (0/34)

Levando-se em conta apenas o limiar 3, foi possível montar a matriz de confusão vista na tabela 9. Ao todo, foram testados 34 valores de log-verossimilhança e o HMM acertou 32. Assim sendo, ele rejeitou erroneamente apenas 2 valores testados.

5.8 Simulação 8: transições determinadas entre os estados do mesmo ambiente de medição C1 (*outdoor*)

A simulação 8 propôs-se a analisar também um conjunto com 4 transições. Posto isso, o ambiente de medição escolhido foi o C1 (*outdoor*). A topologia de Bakis gerou transições pré-definidas de estados, ou seja, definiu-se, de antemão, como ocorreriam as transições entre as localidades RX1 e RX5. Foram estabelecidas 4 transições diretas de estados:

- RX1 para RX2 (transição 1 representado por 1-2).
- RX2 para RX3 (transição 2 representado por 2-3).
- RX3 para RX4 (transição 3 representado por 3-4).
- RX4 para RX5 (transição 4 representado por 4-5).

Essa simulação 8 seguiu os mesmos passos da simulação 7.

O exame dos resultados, com base na análise gráfica, pode ser visto na figura 42. A linha vermelha corresponde aos valores de log-verossimilhança do conjunto de 4 transições. Percebe-se que o limiar 3 apresentou melhor ajuste na identificação das log-verossimilhanças válidas, uma vez que esses valores de log-verossimilhança, em sua maioria, encontram-se à direita do limiar 3.

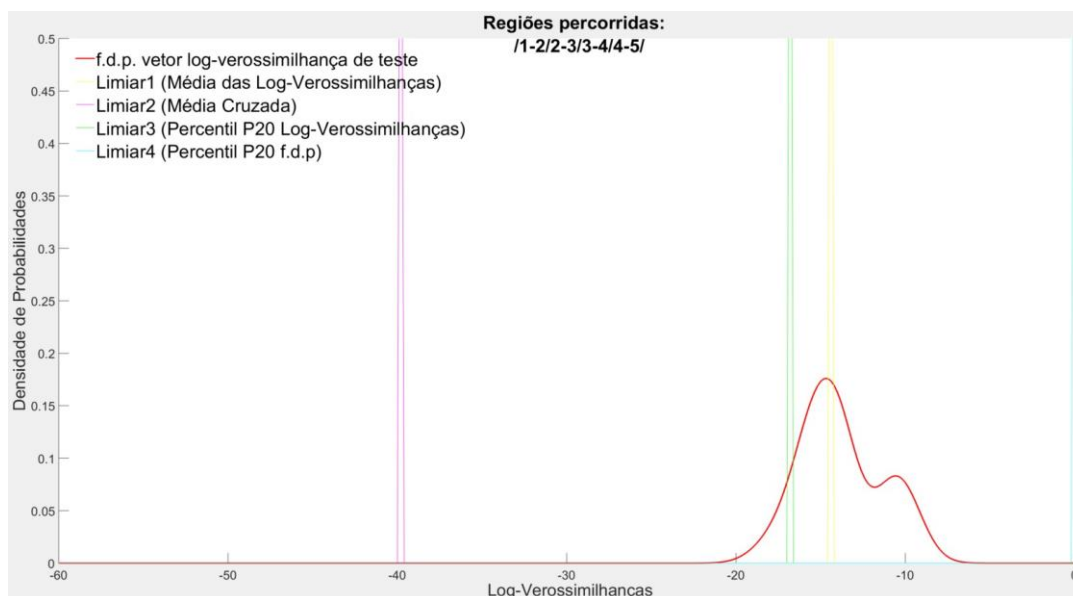


Figura 42- Estimação de f.d.p. para 4 transições do ambiente C1.

Nota-se que não existiram as demais curvas de densidade de probabilidade, visto que a análise aqui é feita com toda a matriz base de características (conjunto total com 4 transições). Além disso, para a matriz de confusão, não existiram dados para as colunas verdadeiro negativo e falso positivo, visto que não houve comparações entre diferentes transições.

Tabela 10 - Matriz de confusão para o conjunto de transições de C1.

Transição de Estados	Verdadeiro Positivo	Falso Negativo	Verdadeiro Negativo	Falso Positivo
1-2 / 2-3 / 3-4/ 4-5	91,18% (31/34)	8,82% (3/34)	0% (0/34)	0% (0/34)

Levando-se em conta apenas o limiar 3, foi possível montar a matriz de confusão vista na tabela 10. Ao todo, foram testados 34 valores de log-verossimilhança e o HMM acertou 31. Assim sendo, ele rejeitou erroneamente apenas 3 valores testados.

5.9 Simulação 9: transições determinadas entre os estados do mesmo ambiente de medição B3 (*outdoor*)

A simulação 9 propôs-se a analisar um conjunto com 5 transições. Posto isso, o ambiente escolhido foi o ambiente de medição B3 (*outdoor*). A topologia de Bakis gerou transições pré-definidas de estados, ou seja, definiu-se, de antemão, como ocorreriam as transições entre as localidades RX1 e RX6. Foram estabelecidas 5 transições diretas de estados:

- RX1 para RX2 (transição 1 representado por 1-2).
- RX2 para RX3 (transição 2 representado por 2-3).
- RX3 para RX4 (transição 3 representado por 3-4).
- RX4 para RX5 (transição 4 representado por 4-5).
- RX5 para RX6 (transição 5 representado por 5-6).

Uma matriz de características de transição do tipo $2 \times 2 \times 36$ foi gerada para cada uma das 5 transições acima mencionadas, agrupando os parâmetros de dispersão temporal τ_m e σ_T das respectivas localidades. Essas matrizes foram reunidas, formando uma matriz base de características do tipo $2 \times 2 \times 36 \times 5$. Metade das amostras da matriz base de características

foi reservada para treinamento do HMM e a outra metade foi destinada para teste e validação. Cabe ressaltar que essa divisão foi feita para cada transição, de modo que tanto a base de treinamento quanto a de teste e validação agrupassem dados das 5 transições. Para inicialização, atribuiu-se valor 6 ao número de estados. Para as misturas gaussianas e número de iterações, fez-se novas varreduras com valores ótimos 2 e 20, respectivamente. Isto posto, para a metade atribuída ao conjunto de teste e validação, seguiu-se com o cálculo dos valores de log-verossimilhança, retirada de *outliers*, divisão das amostras para determinação e teste dos limiares.

O exame dos resultados, com base na análise gráfica, pode ser visto na figura 43. A linha vermelha corresponde aos valores de log-verossimilhança do conjunto de 5 transições. Percebe-se que o limiar 3 é o que apresenta melhor ajuste na identificação das log-verossimilhanças válidas, uma vez que esses valores de log-verossimilhança, em sua maioria, encontram-se à direita do limiar 3. O limiar 2 também apresentou bons ajustes em algumas simulações, não com a mesma frequência ocorrida na simulação 5, mas com frequência suficiente para ser mencionado.

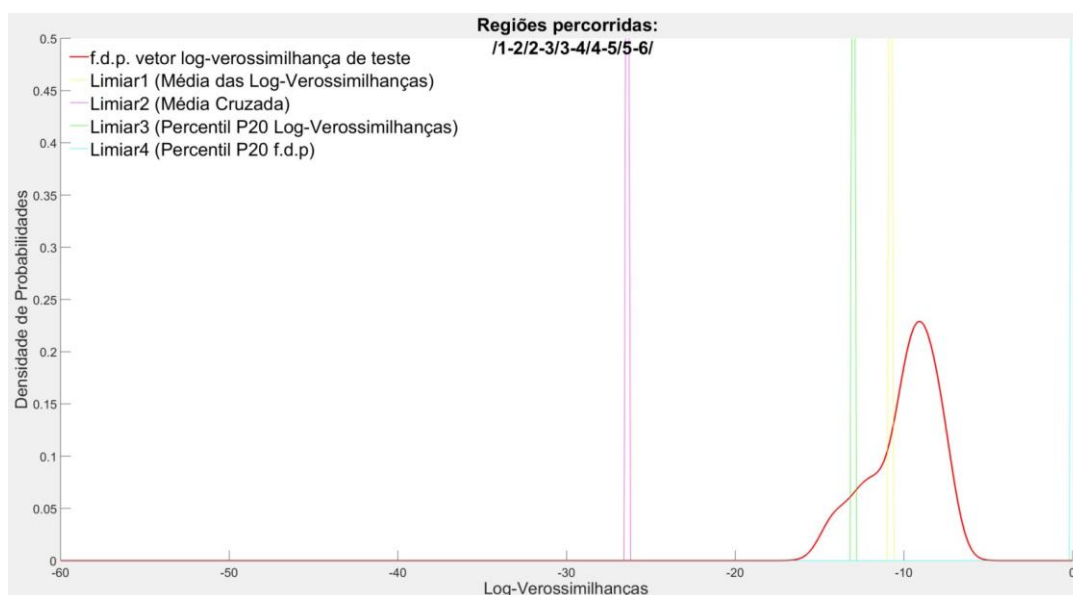


Figura 43- Estimação de f.d.p. para 5 transições do ambiente B3.

Nota-se que não existiram as demais curvas de densidade de probabilidade, visto que a análise aqui é feita com toda a matriz base de características (conjunto total com 5 transições). Além disso, para a matriz de confusão, não existiram dados para as colunas

verdadeiro negativo e falso positivo, visto que não houve comparações entre diferentes transições.

Tabela 11 - Matriz de confusão para o conjunto de transições de B3.

Transição de Estados	Verdadeiro Positivo	Falso Negativo	Verdadeiro Negativo	Falso Positivo
1-2 / 2-3 / 3-4 / 4-5 / 5-6	90,70% (39/43)	9,30% (3/34)	0% (0/34)	0% (0/34)

Levando-se em conta apenas o limiar 3, foi possível montar a matriz de confusão vista na tabela 11. Ao todo, foram testados 43 valores de log-verossimilhança e o HMM acertou 39. Assim sendo, ele rejeitou erroneamente apenas 3 valores testados.

5.10 Simulação 10: transições determinadas entre os estados com grupo de treinamento e grupo de teste de distintos ambientes de medição A5 (*indoor-outdoor*) e A6 (*outdoor-indoor*)

Enquanto as simulações anteriores concentraram esforços nas análises dos seus respectivos ambientes de medição, a simulação 10 englobou dois diferentes ambientes, A5 (*indoor-outdoor*) e A6 (*outdoor-indoor*).

Embora, esses ambientes sejam próximos e, de certo modo, compartilhem áreas comuns, a simulação apresentou resultados relevantes pelo fato de usar um ambiente para treinamento e outro para teste e validação, respectivamente ambientes A5 e A6. Para tal fim, a topologia de Bakis foi aplicada para ambos os ambientes e gerou transições pré-definidas de estados entre as localidades RX1 e RX7. Foram estabelecidas 9 transições:

- RX1 para RX2 (transição 1 representado por 1-2).
- RX2 para RX2 (transição 2 representado por 2-2).
- RX2 para RX3 (transição 3 representado por 2-3).
- RX3 para RX4 (transição 4 representado por 3-4).
- RX4 para RX4 (transição 5 representado por 4-4).
- RX4 para RX5 (transição 6 representado por 4-5).
- RX5 para RX6 (transição 7 representado por 5-6).

- RX6 para RX6 (transição 8 representado por 6-6).
- RX6 para RX7 (transição 9 representado por 6-7).

Logo, para o ambiente A5, destinado ao treinamento do HMM, gerou-se uma matriz de características de transição do tipo $2 \times 2 \times 36$ que agrupou os parâmetros de dispersão temporal τ_m e σ_T , das respectivas localidades de cada uma das 9 transições. Essas matrizes de características de transição foram reunidas em uma matriz do tipo $2 \times 2 \times 36 \times 9$, formando a matriz base de características do ambiente A5.

No lugar de usar metade das amostras da matriz base de características para treinamento e a outra metade para teste e validação, empregou-se completamente essa matriz para treinamento do HMM. Foram 9 transições e, para cada transição, geraram-se 36 combinações aleatórias 2 a 2 (do tipo $2 \times 2 \times 36$) dos parâmetros τ_m e σ_T , de tal maneira que a base montada para treinamento agrupou 324 valores de log-verossimilhanças, 9 x 36 amostras de τ_m e σ_T agrupadas 2 a 2.

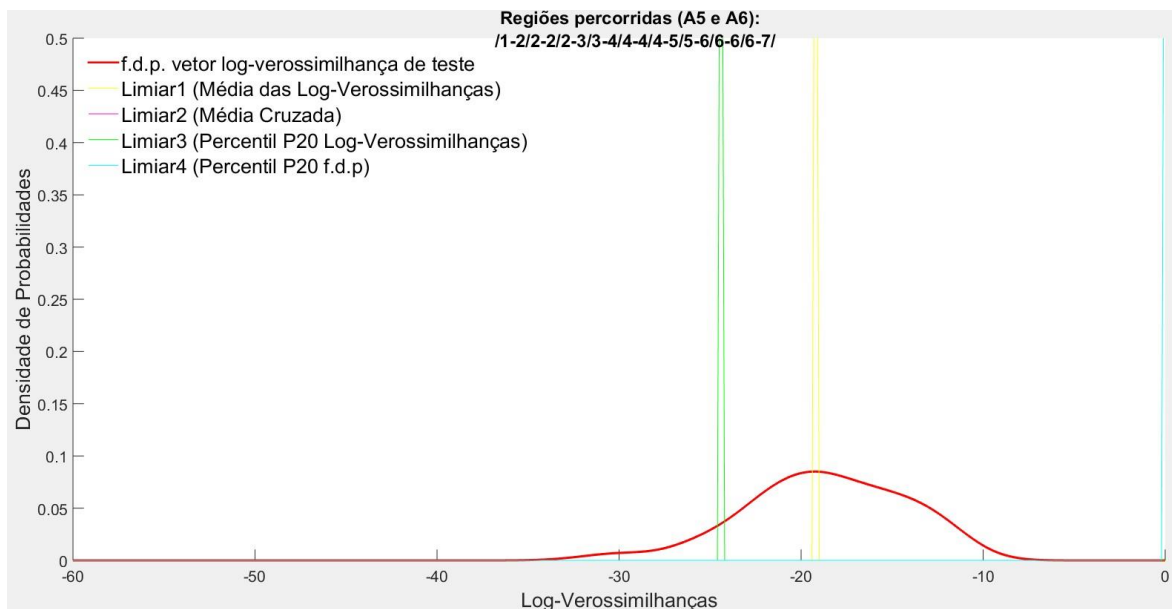


Figura 44- Estimação de f.d.p. para A5 como treinamento e A6 para teste e validação.

Enquanto o ambiente A5 foi dedicado para treinamento do HMM, o ambiente A6 foi empregado para teste e validação. Assim, para o ambiente A6, gerou-se uma matriz base de características, também do tipo $2 \times 2 \times 36 \times 9$, agrupando as matrizes de características de transição, retiraram-se os *outliers* e as log-verossimilhanças foram calculadas. Metade das amostras dessa matriz base de características foi reservada para o grupo de referência

(cálculo dos limiares criados) e a outra metade para o grupo de teste (cálculo dos valores de log-verossimilhanças para estimação de funções de densidade de probabilidade). Vale frisar que a divisão entre grupo de referência e de teste foi feita para cada transição, de tal maneira que tanto o grupo de referência quanto o grupo de teste agrupassem dados de todas as transições.

O exame dos resultados, com base na análise gráfica, pode ser visto na figura 44. A linha vermelha corresponde aos valores de log-verossimilhanças do conjunto total de transições. Percebe-se que o limiar 3 ainda é o que apresenta melhor ajuste na identificação das log-verossimilhanças válidas, uma vez que esses valores de log-verossimilhança, em sua maioria, encontram-se à direita do limiar 3.

Uma vez que a análise aqui é feita com toda a matriz base de características, verifica-se que não existiram as demais curvas de densidade de probabilidade. Também não existiram dados para as colunas verdadeiro negativo e falso positivo na matriz de confusão, uma vez que não houve comparações entre diferentes transições.

Tabela 12 - Matriz de confusão para a simulação com os ambientes A5 e A6.

Transição de Estados	Verdadeiro Positivo	Falso Negativo	Verdadeiro Negativo	Falso Positivo
1-2 / 2-2 / 2-3 / 3-4 / 4-4 / 4-5 / 5-6 / 6-6 / 6-7	90,6% (145/160)	9,4% (15/160)	0% (0/160)	0% (0/160)

Levando-se em conta apenas o limiar 3, é possível montar a matriz de confusão vista na tabela 12. Ao todo, foram testados 160 valores de log-verossimilhanças, o HMM acertou 145 e rejeitou erroneamente apenas 15 valores testados.

5.11 Simulação 11: transições determinadas entre os estados com grupo de treinamento e grupo de teste de distintos ambientes de medição A5 (*outdoor-indoor*) e C1 (*outdoor*)

A simulação 11 seguiu o mesmo raciocínio da simulação 10 e englobou 2 ambientes de medição, A5 e C1, entretanto, a diferença entre as simulações 10 e 11 apresenta-se quando se analisa os ambientes envolvidos. Para a simulação 10, os ambientes

de medição A5 e A6 são próximos, encontram-se no campus da Praia Vermelha e até compartilham áreas comuns. Já para a simulação 11, o ambiente de medição A5 (*indoor-outdoor*) encontra-se no campus da Praia Vermelha, como falado anteriormente, enquanto o campus C1 (exclusivamente *outdoor*) localiza-se no campus do Gragoatá.

Como o ambiente A5 é *indoor-outdoor*, resolveu-se usar apenas as medições *indoor*, de tal forma que a topologia de Bakis gerou transições pré-definidas de estados entre as localidades RX1 e RX5. Foram estabelecidas 6 transições:

- RX1 para RX2 (transição 1 representado por 1-2).
- RX2 para RX2 (transição 2 representado por 2-2).
- RX2 para RX3 (transição 3 representado por 2-3).
- RX3 para RX4 (transição 4 representado por 3-4).
- RX4 para RX4 (transição 5 representado por 4-4).
- RX4 para RX5 (transição 6 representado por 4-5).

Para o ambiente de medição *outdoor* C1, a topologia de Bakis gerou transições pré-definidas de estados entre as localidades RX1 e RX6. Foram estabelecidas 4 transições:

- RX1 para RX2 (transição 1 representado por 1-2).
- RX2 para RX3 (transição 2 representado por 2-3).
- RX3 para RX4 (transição 3 representado por 3-4).
- RX4 para RX5 (transição 4 representado por 4-5).

Para o ambiente A5, destinado ao treinamento do HMM, gerou-se uma matriz de características de transição do tipo $2 \times 2 \times 36$ que agrupou os parâmetros de dispersão temporal τ_m e σ_T de cada uma das 6 transições. Essas matrizes de características de transição foram reunidas em uma matriz do tipo $2 \times 2 \times 36 \times 6$, formando a matriz base de características do ambiente A5. Foram 6 transições e, para cada transição, geraram-se 36 combinações aleatórias 2 a 2 (do tipo $2 \times 2 \times 36$) dos parâmetros τ_m e σ_T , de tal maneira que a base montada para treinamento agrupou 216 valores de log-verossimilhanças, 6×36 amostras de τ_m e σ_T agrupadas 2 a 2.

Para o ambiente C1, destinado ao teste e validação dos limiares, gerou-se uma matriz de características de transição do tipo $2 \times 2 \times 36$ que agrupou os parâmetros de dispersão temporal τ_m e σ_T de cada uma das 4 transições. Essas matrizes de características de transição foram reunidas em uma matriz do tipo $2 \times 2 \times 36 \times 4$, formando a matriz base de

características do ambiente C1. Foram 4 transições e, para cada transição, geraram-se 36 combinações aleatórias 2 a 2 (do tipo 2x2x36) dos parâmetros τ_m e σ_T , de tal maneira que a base de teste e validação agrupou 144 valores de log-verossimilhança, 4 x 36 amostras de τ_m e σ_T agrupadas 2 a 2.

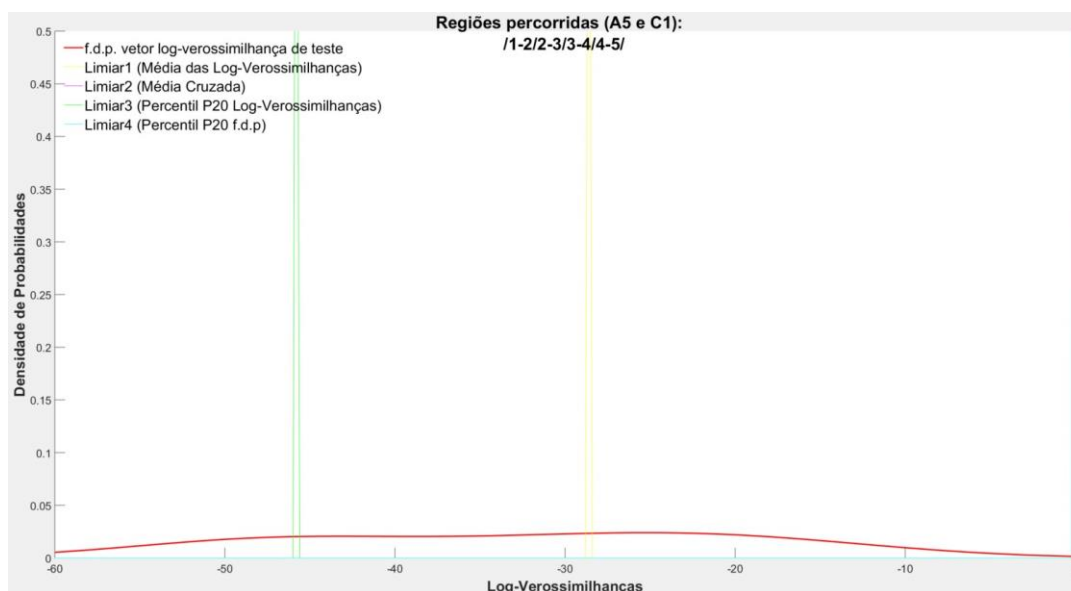


Figura 45- Estimação de f.d.p. para A5 como treinamento e C1 para teste e validação.

O exame dos resultados, com base na análise gráfica, pode ser visto na figura 45. A linha vermelha corresponde aos valores de log-verossimilhanças do conjunto total de transições.

Tabela 13 - Matriz de confusão para a simulação com ambientes A5 e C1.

Transição de Estados	Verdadeiro Positivo	Falso Negativo	Verdadeiro Negativo	Falso Positivo
1-2 / 2-3 / 3-4 / 4-5	71,43% (50/70)	28,57% (20/70)	0% (0/70)	0% (0/70)

Uma vez que a análise aqui é feita com toda a matriz base de características, verifica-se que não existiram as demais curvas de densidade de probabilidade. Também não existiram dados para as colunas verdadeiro negativo e falso positivo na matriz de confusão, uma vez que não houve comparações entre diferentes transições.

Levando-se em conta apenas o limiar 3, é possível montar a matriz de confusão vista na tabela 13. Ao todo, foram testados 70 valores de log-verossimilhança, o HMM acertou 50 e rejeitou erroneamente 20 valores testados.

Perceba que o objetivo dessa simulação 11 era justamente verificar se o HMM erraria mais do que errou nas demais simulações, visto que ele estava usando como base de treinamento um ambiente de medição completamente diferente. Os resultados encontrados corroboram o esperado e o HMM acertou uma porcentagem bem abaixo das simulações anteriores.

5.12 Resultados

Os sistemas de comunicações móveis estão em contínua evolução, haja vista a atual quinta geração dos sistemas de comunicações móveis celulares, IMT-2020. Ao passo que esses sistemas evoluem para suprir as demandas e necessidades dos consumidores, análises mais apuradas dos meios de propagação tornam-se vitais para o apropriado entendimento dos mecanismos e fenômenos inerentes à propagação, bem como do adequado dimensionamento do CRM para mitigar as limitações impostas pelos meios de propagação e os custos elevados de eventuais medições de campo.

Tendo em consideração o exposto acima, essa dissertação objetivou a caracterização do CRM fazendo uso do método estatístico HMM. De forma a estruturar os resultados alcançados, todo o processo foi aqui dividido em três etapas, tendo cada uma suas considerações e apontamentos. A primeira etapa foi a definição dos ambientes de medição, adoção da técnica de sondagem do CRM, realização das medições oficiais em ambientes específicos. A segunda etapa correspondeu ao uso de malhas quadrangulares 6 por 6, a obtenção das estatísticas de pequena escala do CRM para cada posição de recepção, ao pré-processamento dos dados medidos e a adequação desses dados pré-processados ao *programa* Matlab. A terceira etapa correspondeu, enfim, à caracterização do CRM com uso do HMM, onde se adotou a topologia de Bakis para as transições entre estados, os agrupamentos dos dados gerados pelas transições foram divididos em bases de treinamento para o HMM e bases de teste para definições de limiares e estimações de f.d.p. e, por fim, os resultados foram obtidos através de análises gráficas e matrizes de confusão.

5.12.1 Primeira etapa: medições realizadas

Os ambientes da campanha de medições foram escolhidos de forma que a presente dissertação prosseguisse com resultados adicionais aos resultados, até então, encontrados por Ribeiro [14]. Dessa forma, os ambientes adotados foram A5 (indoor-outdoor), A6 (outdoor-indoor), A7 (indoor-outdoor), A8 (outdoor-indoor), B3 (outdoor), B4 (outdoor), C1 (outdoor) e C4 (outdoor).

A técnica adotada para sondagem do CRM banda larga foi a técnica de varredura em frequência (*swept frequency – chirp*). Ela mostrou-se como uma opção viável, uma vez que foram adotadas moderações como calibração dos equipamentos, grande sincronismo entre transmissor e receptor, medições em ambientes com curtas distâncias e rápidos intervalos de varredura de sinal no CRM. Posto essas restrições, a técnica de varredura em frequência apresentou medições adequadas, parâmetros como retardo máximo ($\tau_{\text{máx}}$) e resolução temporal ($\Delta\tau$) tiveram valores dentro do esperado.

Além disso, medições realizadas em períodos de baixo movimento e a completa imobilidade das pessoas envolvidas no processo de medições, durante os intervalos de varredura de sinal no canal, garantiram a estacionariedade do CRM de forma satisfatória.

Para cada varredura de 696 ms realizada em cada posição fixa da malha quadrangular 6 por 6, a técnica de varredura em frequência permitiu a aquisição com sucesso da função de transferência de tempo variante $T(f, t)$ do CRM. Como, praticamente, havia uma estacionariedade temporal para cada curto intervalo de varredura do sinal, $T(f, t)$ estava composta apenas de amostras discretas de amplitudes e fases referentes às respostas em frequência do CRM dentro da banda sondada.

5.12.2 Segunda etapa: pré-processamento dos dados

Sendo o canal considerado como um filtro linear, os registros de entrada e saída caracterizaram diretamente a função de transferência de tempo variante $T(f, t)$. Em posse dessa função, usou-se, com sucesso, a função janela Blackman-Harris de 3 termos para minimizar o vazamento espectral provocado pela discretização,

Aplicou-se a IDFT a $T(f, t)$ para obtenção dos perfis de potência de retardo $P_h(\tau)$, visto que o canal foi considerada QWSSUS. Para os perfis obtidos, aplicou-se a técnica

CFAR para limpeza de ruído [39], evidenciando, com sucesso, apenas os perfis válidos, ou multipercursos válidos para a análise. De forma geral, o número de multipercursos está ligado a características de LoS ou NLoS e a distância entre transmissor e receptor. Maiores distâncias e a falta de visada direta tendem a aumentar os parâmetros de dispersão temporal do canal. Os ambientes A5 (*indoor-outdoor*) e A6 (*outdoor-indoor*) apresentaram, aproximadamente, os mesmos valores para os parâmetros de dispersão temporal do canal. Já o ambiente C1 (*outdoor*) apresentou maiores valores para os parâmetros de dispersão temporal, dada diferença de ambiente e a presença de folhagens.

As malhas quadrangulares de 36 posições permitiram a aquisição de pontos de medições descorrelacionados em cada localidade. O levantamento de estatísticas de pequena escala foi aplicado em cada ponto para obtenção dos principais parâmetros de dispersão temporal do CRM: retardo médio (τ_m) e espalhamento de retardo (σ_T). Assim, para cada ponto, esses parâmetros foram agrupados em matrizes bidimensionais do tipo 36 x 2 e foram chamadas de matrizes de características por localidade. Essas matrizes já foram criadas de forma a se adequarem ao funcionamento do *toolkit* do HMM para Matlab.

5.12.3 Terceira etapa: caracterização do CRM com uso do HMM

As matrizes bidimensionais de características por localidade foram agrupadas, reunindo aleatoriamente dois a dois os 36 parâmetros de τ_m e σ_T e formaram matrizes tridimensionais do tipo 2 x 2 x 36, chamadas de matrizes de características de transição. Quando todas essas matrizes tridimensionais foram agrupadas, conforme topologia de Bakis adotada para as transições entre estados, as matrizes base de características do CRM foram criadas com sucesso.

As matrizes base de características foram divididas em bases de treinamento para o HMM e bases de teste para definições de limiares e estimações de f.d.p. Por fim, os resultados foram obtidos através de análises gráficas e matrizes de confusão.

A tabela 14 agrupa características das simulações, para que, de forma resumida, seja possível analisar os resultados finais do trabalho efetuado. Esses resultados foram promissores para os diferentes ambientes analisados. Trabalhando com parâmetros temporais do CRM em faixa larga, obtidos através de testes de campo em diferentes ambientes (*indoor-outdoor*, *outdoor-indoor* e *outdoor*), as aplicações do método HMM nas

simulações de transições entre parâmetros de dispersão temporal (τ_m e σ_T) legitimaram o uso do HMM para análise e caracterização do CRM com diferentes quantidades de transições de estados e em diferentes ambientes.

Tabela 14 – Agrupamento de características e resultados das simulações.

Nº da simulação	Limiar usado	Tipo de ambiente	Base treino	Base teste	Nº de transições	Verdadeiro Positivo	Falso Negativo
4	3	<i>indoor-outdoor</i>	A7	A7	3	92,00%	8,00%
5	3	<i>outdoor-indoor</i>	A8	A8	3	92,00%	8,00%
6	3	<i>outdoor</i>	C4	C4	3	96,00%	4,00%
7	3	<i>outdoor</i>	B4	B4	4	94,12%	5,88%
8	3	<i>outdoor</i>	C1	C1	4	91,18%	8,82%
9	3	<i>outdoor</i>	B3	B3	5	90,70%	9,30%
2	3	<i>indoor-outdoor</i>	A5	A5	9	90,10%	8,90%
3	3	<i>outdoor-indoor</i>	A6	A6	9	90,10%	8,90%
10	3	A5 <i>indoor-outdoor</i> A6 <i>outdoor-indoor</i>	A5	A6	A5: 9 A6: 9	90,60%	9,40%
11	3	A5 <i>indoor-outdoor</i> C1 <i>outdoor</i>	A5	C1	A5: 6 C1: 4	71,43%	28,57%

Nota-se que, com o aumento do número de transições, o percentual de verdadeiro positivo diminuiu. Isso pode ser explicado pela quantidade limitada de dados disponíveis para análise. Ou seja, seriam necessários mais dados para analisar um conjunto maior de possibilidades de transições e, dado a limitação da quantidade de dados disponíveis (dados das medições realizadas), o percentual de verdadeiro positivo diminuiu quando se analisou um conjunto maior de transições.

Capítulo 6 - Conclusões e trabalhos futuros

6.1 Conclusões

Após todas as simulações e resultados encontrados, podemos validar o método proposto. Destacamos a seguir as principais conclusões do trabalho.

A partir das análises gráficas das f.d.p. versus limiares calculados, foi possível verificar que o limiar 3, percentil P20 das log-verossimilhanças, foi o mais efetivo na identificação das log-verossimilhanças válidas em todas as simulações. Ou seja, foi o limiar onde a grande maioria das log-verossimilhanças do grupo de testado (grupo de log-verossimilhanças válidas) se agrupou acima (ou a direita) dele. É válido comentar que o limiar 2, média cruzada, também apresentou bom ajuste para as simulações 5 e 9 e que os demais limiares não apresentaram bons ajustes para nenhuma simulação.

A partir das análises das matrizes de confusão, foi possível classificar o posicionamento das log-verossimilhanças em relação ao limiar 3, limiar usado como referência, uma vez que ele foi o mais efetivo na identificação das log-verossimilhanças válidas.

A simulação 1, ambiente de medição A5 *indoor-outdoor*, com análises individuais das transições aleatórias entre estados (tabela 3), mostrou que o HMM acertou a grande maioria dos valores de todas as transições de estados correntes (coluna verdadeiro positivo) e errou poucos valores das transições desses estados correntes (coluna falso negativo). Para a comparação das transições de estados correntes com os demais estados, o HMM apresentou resultados satisfatórios para as transições de estados 1-1, 1-3, 3-4 e 7-7 (valores elevados de verdadeiro negativo e valores baixos de falso positivo) e resultados medianos para as transições de estados 4-5, 5-5 e 5-7. A justificativa para os valores medianos entre verdadeiro negativo (o HMM acertou que eram valores de outras transições e não da transição analisada) e falso positivo (o HMM errou ao considerar que os valores pertenciam à transição analisada e não a outras transições) pode recair sobre o conjunto de dois fatores: as transições de estados 4-5, 5-5 e 5-7 estavam localizadas logo após a passagem do ambiente *indoor* para o ambiente *outdoor* e, como cada transição foi avaliada individualmente, havia muito mais amostras das outras transições do que amostras da própria transição analisada, ou seja, o conjunto de log-verossimilhanças de outras transições era maior e mais diversificado.

Para o conjunto de simulações de 2 a 11, veja que as análises das matrizes de confusão foram feitas com todas as respectivas matrizes base de características, de forma que não existiram comparações entre transições de estados correntes e transições dos demais estados, também não existiram dados para as colunas verdadeiro negativo e falso positivo e as verificações concentraram-se nas colunas verdadeiro positivo e falso negativo. Além disso, percebe também que existem simulações com diferentes quantidades de transições de estados. Dessa forma, o HMM apresentou resultados satisfatórios (acertos acima de 90%) para todas as simulações efetuadas, exceto para a simulação 11 (tópico 5.11).

Para o conjunto de simulações de 2 a 10, com ambientes *indoor-outdoor*, *outdoor-indoor* e *outdoor*, o HMM acertou (sempre acima de 90%) que os valores testados pertenciam ao conjunto de transições analisadas, ou seja, eram valores de log-verossimilhanças válidas do ambiente analisado.

Para a simulação 11, com grupo de treinamento pertencente ao ambiente A5 (*outdoor-indoor*) e grupo de teste pertencente ao ambiente C1 (*outdoor*), o HMM acertou apenas que 71,43% dos valores testados pertenciam ao conjunto de transições analisadas. Esse baixo desempenho era esperado. A simulação 11 tinha como objetivos: mostrar a existência de duas bases de dados diferentes (base de treinamento e base de teste), mostrar que o HMM realizou treinamento com a base de dados do ambiente A5 (apesar desse ambiente ser *indoor-outdoor*, usou-se, para treinamento, apenas as localidades *indoor*), mostrar que o HMM realizou testes com a base de dados do ambiente C1 (*outdoor*), e, uma vez que esses ambientes eram diferentes, o HMM teria baixo desempenho.

É justo fazer uma ressalva para o ambiente *outdoor* utilizado nessa dissertação. Aqui ele é considerado como um ambiente externo, mas de curta distância, distinguindo-se de outras literaturas que tratam o ambiente *outdoor* como um ambiente externo que cobre grandes distâncias.

Em vista do exposto acima, conclui-se que o método estatístico HMM ratifica as estatísticas dos parâmetros de retardo médio (τ_m) e espalhamento de retardo (σ_T) de forma a validar o uso do HMM para caracterização do CRM com diferentes quantidades de transições, diferentes quantidades de estados e em diferentes ambientes (*indoor-outdoor*, *outdoor-indoor* e *outdoor*).

O HMM é usado na transição entre fonemas para a modelagem da voz como um sinal variante no tempo. O presente trabalho, para além do conhecimento da potência na

recepção, interessou-se em como os parâmetros do CRM estavam variando nas transições entre localidades de recepção. Logo, o fato do perfil de retardos ser variável com o tempo justifica o porquê do método HMM apresentar bons resultados.

6.2 Sugestões para trabalhos futuros

São sugeridas as seguintes linhas de estudo para trabalhos futuros:

- Para trabalhos futuros, avistam-se possibilidades de estudos comparativos com mais quantidades de transições de estados, bem como com diferentes cenários de medição do CRM como em ambientes densamente urbanos, urbanos, suburbanos, residenciais, rurais e veiculares.
- Busca-se uma validação da aplicação do método HMM para modelagem do CRM banda larga, usando parâmetros de dispersão temporal do canal. Uma vez que se tomou como inspiração a identificação de um locutor para uma caracterização local (em um ambiente) do CRM, pode-se pensar na identificação de uma palavra ou frase, independente do locutor, como inspiração para um modelo mais generalizado de caracterização do CRM com uso do HMM.
- Uma possível aplicação está relacionada à adequação da transmissão dado o tipo de ambiente. Um dispositivo móvel celular realiza medições de potência em intervalos regulares. Para cada intervalo, usando HMM, classifica-se o tipo de ambiente onde o dispositivo móvel celular se encontra e, dado o tipo de ambiente, a estação base define um modo de transmissão. Em um novo intervalo, caso o HMM indique mudança na classificação do ambiente, a estação base poderá ajustar parâmetros de transmissão que possibilite a adequada comunicação com o dispositivo móvel celular. Outro exemplo para uma possível aplicação, em redes convergentes, o próprio dispositivo móvel celular classifica o ambiente como *indoor* e troca a rede de telefonia móvel por uma rede Wi-Fi.

Bibliografia

- [1] MASSIE, W. W., UNDERHILL, C. R. "Wireless Telegraphy & Telephony: The Future of Wireless Art", p. 67-71., Nova Iorque, Estados Unidos: D.Van Nostrand Inc., 1908.
- [2] HAYKIN, S., MOHER, M. "Sistemas Modernos de Comunicações Wireless", 1ª ed., São Paulo, Brasil: Bookman Companhia Editora Ltda., 2008.
- [3] GSMA Intelligence Report, "GSMA The Mobile Economy 2023," 27 2 2023. [Online]. Available: <https://www.gsma.com/mobileeconomy/>. [Acesso em 16 3 2023].
- [4] 3GPP Release 6, Technical Report (TR) 21.101 V6.10.0 (2009-12), "3gpp.org," 2012. [Online]. Available: <https://www.3gpp.org/specifications-technologies/releases/release-6>. [Acesso em 15 3 2023].
- [5] 3GPP Release 15, Technical Report (TR) 21.915 V15.0.0 (2019-09), "3gpp.org," 9 2019. [Online]. Available: <https://www.3gpp.org/specifications-technologies/releases/release-15>. [Acesso em 15 3 2023].
- [6] 3GPP Release 16, Technical Report (TR) 21.916 V16.1.0 (2022-01), "3gpp.org," 1 2022. [Online]. Available: <https://www.3gpp.org/specifications-technologies/releases/release-16>. [Acesso em 15 3 2023].
- [7] 3GPP Release 17, Technical Report (TR) 21.917 V0.2.0 (2022-01), "3gpp.org," 1 2022. [Online]. Available: <https://www.3gpp.org/specifications-technologies/releases/release-17>. [Acesso em 15 03 2023].
- [8] SALEH, A. A. M., VALENZUELA, R. A. "A statistical model for indoor multipath propagation", IEEE J. Select. Areas Commun., v. 5, pp. 128–137, Fevereiro 1987.
- [9] COELHO, F. V. S., CATALDO, E., MATOS, L. J. "Simulation of Indoor-Outdoor and Outdoor-Indoor Radio Channel in 960-1700 MHz Band". IEEE Latin America Transactions, v. 15, p. 1400-1406, 2017.
- [10] HORWARD, S., PAHLAVAN, K. "Autoregressive Modeling of Wide-Band Indoor Radio Propagation", IEEE Transaction on Commun., v. 40, pp. 1540-1552, Setembro 1992.
- [11] SILVA, F. A. L., MATOS, L. J., CATALDO, E. "Modelo auto-regressivo aplicado a um canal rádio de banda ultralarga". XVI Simpósio Brasileiro de Telecomunicações – SBrT'08, 02-05 de Setembro de 2008, Rio de Janeiro, RJ.
- [12] IWAI, H., KARASAWA, Y. "Wideband Propagation Model for the Analysis of the Effect of the Multipath Fading on the Near-Far Problem in CDMA Mobile Radio Systems", IEICE, Trans. Comm., Vol E 76-B, pp-103-112, fev 1993.
- [13] ALMEIDA, R. V. S. Almeida, MATOS, L. J. "Emprego da técnica de geração e desaparecimento de raios na simulação de canal UWB em ambiente indoor". XXVII Simpósio Brasileiro de Telecomunicações – SBrT 2009, de 29 de Setembro a 02 de Outubro de 2009, Blumenau.
- [14] RIBEIRO, L. G., MATOS, L. J., CATALDO, E. "Aplicação de modelos ocultos de Markov na modelagem do canal rádio móvel". MOMAG2020, 2020, Niterói. 19°. Simpósio Brasileiro de Micro-Ondas e Optoeletrônica/14°. Congresso Brasileiro de Eletromagnetismo. São Caet.
- [15] JELINEK, F., "Continuous Speech Recognition by Statistical Methods", Proc. IEEE,

- Vol. 64, pp. 532-556, April 1976.
- [16] RABINER, L. R., "A Tutorial on HMM and Select App in Speech Recognition", Proc. IEEE, Vol. 77, n. 2, Feb. 1989.
- [17] RABINER, L. R., JUANG, B. H., "An Introduction to Hidden Markov Models". IEEE ASSP Magazine, 3, 4-16.
- [18] PORITZ, A. B., "Hidden Markov Model: A Guided Tour", Proc. of the IEEE Int. Conf. on in Acoustics, Speech, and Signal Processing, pp. 7- 13, 1988.
- [19] F. M. S. C. F. C. E. SILVEIRA, "Utilização da técnica MFCC em conjunto com os parâmetros extraídos do sinal glotal para melhorar o desempenho de um sistema de verificação de locutor", *XXXV Simpósio Brasileiro de Telecomunicações, São Pedro, SP.*, 3 a 6 9 2017.
- [20] CATALDO, E., SOIZE, C., "Voice signals produced with jitter through a stochastic one-mass mechanical model," *Journal of Voice*, vol. 31, no 1, pp. 111e9–111e18, 2017.
- [21] L. G. RIBEIRO, "ppgeet.uff.br," 4 11 2019. [Online]. Available: <http://www.ppgeet.uff.br/site/wp-content/uploads/2021/02/Leonardo-Guimaraes.pdf>.
- [22] C. E. SHANNON, "A Mathematical Theory of Communication," *The Bell System Technical Journal*, vol. 27, n. 3, pp. 379–423, 623–656, 1948.
- [23] STEELE, R., "Mobile Radio Communications", Piscataway, Nova Jersey, Estados Unidos: IEEE Press, 1994.
- [24] PARSONS, J. D., "The Mobile Radio Propagation Channel", 2ª ed., Chichester, Inglaterra: John Wiley & Sons Inc., 2000.
- [25] RAPPAPORT, T. S., "Wireless Communications Principles and Practice". 2ª ed., Nova Jersey, Estados Unidos: Prentice Hall Professional Technical Reference, 2002.
- [26] LEON-GARCIA, A., "Probability, Statistics, and Random Processes For Electrical Engineering", 3ª ed., Nova Jersey, Estados Unidos: Pearson Education Inc., 2008.
- [27] ITU-R Radiocommunication Sector of ITU, "P.1411 : Propagation data and prediction methods for the planning of short-range outdoor radiocommunication systems and radio local area networks in the frequency range 300 MHz to 100 GHz," 9 2021. [Online]. Available: https://www.itu.int/dms_pubrec/itu-r/rec/p/R-REC-P.1411-11-202109-I!!PDF-E.pdf. [Acesso em 15 3 2023].
- [28] LEE, W. C. Y., "Mobile Communications Design Fundamentals", 2ª ed., Nova Iorque, Estados Unidos: John Wiley & Sons Inc., 1993., [Online].
- [29] ZADEH, L. A., "Frequency analysis of variable networks". Proceedings I.R.E., v. 38, p. 291-299, 1950.
- [30] KAILATH, T., "Sampling models for linear time-variant filters". Report 352, MIT Research Lab of Electronics, Cambridge, MA, 1959.
- [31] BELLO, P. A., "Characterization of randomly time-variant linear channels". IEEE Transactions on Communication Systems, v. CS-11, p. 360-393, Dez. 1963.
- [32] BELLO, P. A., "Time-frequency duality". IEEE Transactions, IT10(1), 18-33.
- [33] MERZBACH, U. C., BOYER, C. B., "A History of Mathematics", 3ª ed., Nova Iorque, Estados Unidos: John Wiley & Sons Inc., 2011.
- [34] JELINEK, F., "Statistical Methods for Speech Recognition", Estados Unidos: The MIT Press, 1998.
- [35] COELHO, J. P., PINHO, T. M., BOAVENTURA-CUNHA, J., "Hidden Markov

- Models: Theory and Implementation using Matlab®", Flórida, Estados Unidos: CRC Press Inc., 2019.
- [36] BISHOP, C. M., "Pattern Recognition and Machine Learning", Nova Iorque, Estados Unidos: Springer, 2006.
- [37] HAYKIN, S. "Redes Neurais: Princípios e Prática", 2ª ed., São Paulo, Brasil: Bookman Companhia Editora Ltda., 2000.
- [38] C. F. SOUZA, "ppgeet.uff.br", 28 7 2006. [Online]. Available: http://www.ppgeet.uff.br/site/wp-content/uploads/2021/01/Christiano_Freitas_De_Souza.pdf.
- [39] L. J. Matos, B. S. MARINHO, "A comparison of the delay spread obtained with different power delay profiles de-noising techniques", *Engevista*, vol. 13, n. 2, pp. 129-133, 2011.
- [40] M. CHEN, "MATLAB Central File Exchange, Hidden Markov Model Toolbox (HMM)," 16 3 2023. [Online]. Available: <https://www.mathworks.com/matlabcentral/fileexchange/55866-hidden-markov-model-toolbox-hmm>. [Acesso em 15 3 2023].

Anexo A – Programa em Matlab: pré-processamento dos dados

Segue o código do programa em Matlab usado para pré-processamento dos dados.

```
% Universidade Federal Fluminense (UFF)
% Pós-Graduação em Engenharia Elétrica e de Telecomunicações (PPGEET)
% Tese: Aplicação de HMM na modelagem de CRM - Leonardo Ribeiro (2019)
%
% clear all;
clear all % Remove variáveis da área de trabalho. Libera memória do sistema.
close all % Exclui todas as figuras cujas alças não estejam ocultas
clc; % Limpa a janela de comando (Clear Command Window)
%
B = 750; % [MHz] - Banda de frequência sondada = 1710 - 960 MHz
pontos = 1601; % Varredura de B com 1601 amostras de frequência em cada ponto
Nmeio = (pontos+1)/2; % 801 pontos/amostras de medição
deltaF = B/(pontos-1); % [MHz] Resolução na frequência (intervalo entre as amostras) = 0,46875MHz
%
RetardoMax = 10^3/deltaF; % [nseg] -> Duração de 1 perfil = 2,1333 x 10^3 nseg
% RetardoMax = tempo máx entre 2 medições em um mesmo ponto
% deltaF em [MHz]. RetardoMax em [micros] = 10^-6
% RetardoMax em [nseg] = 10^-9 -> RetardoMax = (1x10^3)/(deltaFx10^3)
% [Christiano 2006 - Pág 65] Representa o maior valor de retardo que se
% consegue medir, e está associado ao sinal resultante do somatório dos
% multipercursos, que resultam na maior distância entre os pontos TX e RX.
%
deltaT = 10^3/B; % [nseg] -> Resolução temporal = 1,3333
% Mesmo raciocínio. x10^3 para alterar a unidade para [nseg]
% % % % % % % % CHAMANDO OS ARQUIVOS GRAVADOS DA MEDIÇÃO % % % % % % % %
%
numlocal = input('Entre com o nº do LOCAL DE MEDIÇÃO desejado (RX que pode variar entre 1 e 7): ');
% Nº do LOCAL DE MEDIÇÃO representa cada um dos 7 pontos de medição (RX1 ~ RX7)
% A medição em cada LOCAL DE MEDIÇÃO foi feita com GRIDS de 36 posições.
% 36 funções (1 função para ponto) foram obtidas para cada GRID
% (cada GRID é uma matriz 6x6 com 15cm de afastamento entre os pontos)
% Freq banda superior = 1710MHz >> lambda = 17,54cm >> 0,38lambda = 6,7cm
% Freq banda inferior = 960MHz >> lambda = 31,25cm >> 0,38lambda = 11,9cm
% 1a correlação zero entre sinais (sinais descorrelacionados) = 0,38 lambda
% Na prática, utiliza-se 0,50 lambda (aqui usou 0,38 lambda)
% Esse método possibilita a obtenção de estatísticas do canal, já que em
% cada ponto do grid a RX do sinal está descorrelacionada dos demais
%
% Loop (kp) para incrementar até o nº máximo de GRIDS num local(numlocal)
%
Amp = zeros([36 1601]); % INSERI MATRIZ DE ZEROS
T = zeros([36 1601]); % INSERI MATRIZ DE ZEROS
Vetor = zeros([1 1601]); % INSERI MATRIZ DE ZEROS
aux = zeros([36 1601]); % INSERI MATRIZ DE ZEROS
%
No_Perfis = 36; % 36 pontos de medição em cada LOCAL DE MEDIÇÃO
%
for kp = numlocal; % kp = Nº do LOCAL DE MEDIÇÃO
%No_Perfis = 36; % Grid 36 pnts de medição em cada LOCAL DE MEDIÇÃO
pp = int2str(kp); % Conversão p/ string p/ criar nome 'ponto1_1.mat'
nome = ['ponto',pp]; % Variável 'nome' recebe string ponto + pp
% Se pp = 1 -> nome = 'ponto1'
```

```

% Loop para incrementar até No_Perfis medidos em cada GRID (ki)
%
for ki = 1:No_Perfis; % ki é o nº do perfil ( = ponto do grid )
    rr = int2str(ki); % Conversão para string
    arquivo = [nome,'_',rr];% Criação do nome do arquivo 'ponto1_1'
    load (arquivo); % Carrega 'arquivo' correspondente para o MatLab
    % Exemplo: para kp = 1, então carrega 'arquivo' com nomes 'ponto1_1' até 'ponto1_36'
    % Cada arquivo 'ponto1_1' até 'ponto1_36' é uma matriz M de 2x1601
    % 1a linha corresponde as amplitudes RX dos sinais (M 1x1601)
    % 2a linha corresponde as fases RX dos sinais (M 2x1601)
    Amp2(1,:) = varre(1,:); % [dB]. Gera array com 1601 medições
    Fase2(1,:) = varre(2,:); % Em graus. 2 Linha array com 1601 medições
    % Varre(1,:) = Retorna a 1a linha, Varre(2,:) = Retorna a 2a linha
    % Carrega no MATLAB as medições escolhidas com 'load (arquivo)'.
    % Valores armazenados nas variáveis Amp2 e Fase2.
    % Amp2 = Array com medições de potência em dB.
    % Fase2 = Array com medições de Fase em graus.
    % Amp2 e Fase2 geram uma matriz 2x1601 com a variável Varre
    %
    ponto1_1 = varre;
    % clear varre; % Remove a variável 'varre' da área de trabalho,
    % libera memória do sistema. 'varre' armazena Amp2 e Fase2
    %
    % Parâmetros Estatísticos da Sondagem
    %
    retardo = 0:deltaT:RetardoMax/2; % [nseg] 1/2 perfil = 801 pontos
    % x:y:z -> Início em x. Fim em z. De x até z com incremento de y.
    % De 0 até [(2,1333x10^3 nseg)/2] com incremento de 1,3333
    %
    % Composição dos dados em números complexos
    %
    Amp(ki,:) = 10.^(Amp2(1,:)/20); % [V] Amplitude em volts,
    % pois medição(Amp2) está em dB
    % Amp = Matrix 36x1601. Para o ponto de Medição RX1, a variável Amp
    % está mostrando as 1601 medições de amplitude [V] de cada ponto do
    % Grid. 1601 medições em cada um dos 36 pontos do Grid.
    %
    T(ki,:) = Amp(ki,:).*(cos(Fase2(1,:)*pi/180)+i*Amp(ki,:).*...
    sin(Fase2(1,:)*pi/180)); % T(f,t) -> Pág 36
    % pi/180 = fase em radianos
    % Pois MATLAB calcula o COS e SEN de rad e não de graus
    %
    % "O tempo t foi mantido fixo, correspondendo a uma 'varredura'.
    % A T(f,t) é composta apenas de amostras em frequência discretas
    % dentro da banda sondada B" -> Pág 36 Tese Leonardo
    % Relação de VOLTS x FREQ
    %
    % Transformada De Fourier de um sinal finito da margem para
    % espúrios (parte ã real) no domínio da frequência (conhecido
    % como Vazamento de Frequência Espectral - Spectral Leakage).
    %
    % Aplicando a Janela de Blackman Harris para reduzir espúrios da
    % resposta do canal no domínio dos retardos (associados a
    % intervalos finitos de observação no domínio da frequência),
    % antes da Transformada de Fourier os dados discretos são aplicados
    % a uma função janela para redução dos lóbulos secundários.
    % Janela de Blackman Harris possui uma das melhores performances.
    %
    a0 = 0.42323; % Coeficientes 1 Janela Blackman Harris c/ 3 termos
    a1 = 0.49755; % Coeficientes 2 Janela Blackman Harris c/ 3 termos

```

```

a2 = 0.07922; % Coeficientes 3 Janela Blackman Harris c/ 3 termos
% Vetor = blackmanharris(pontos,'periodic'); % Array/matriz coluna
% wvtool(Vetor)
% PBH = Vetor'; % Gera array/matriz linha a partir do array Vetor
%
% Função blackmanharris() % Minimum 4-term Blackman-Harris window
% Janela Blackman Harris c/ 4 termos (a0, a1, a2 e a3)
% a0 = 0.35875, a1 = 0.48829, a2 = 0.14128, a3 = 0.01168
%
for n = 0:(pontos-1)
    Vetor(n+1) = a0-a1*cos(2*pi*n/pontos)+a2*cos(2*pi*2*n/pontos);
end
PBH = Vetor'; % Gera array com 1601 medições armazenadas em Vetor
clear Vetor % Remove a variável 'Vetor' da área de trabalho, libera memória do sistema
%
% % % Aplicando a IFFT (transformada inversa) a T(f,t) para chegar a h(t,tau) - Pág18 % % %
aux(ki,:) = ifft( (T(ki,:)).*PBH,pontos ); % h(tau)
% h(tau) - relação de volts x retardo -> 1601 pontos
PPRR = abs(aux); % |h(tau)|/1601 pontos
Pot = PPRR.^2; % Ph(tau)/Relação de WATTS/ 1601 pontos
Potdb = 10.*log10(Pot); % Perfil de Potencia relativo Ph(tau), em dB/ 1601 pontos
clear Pot % Remove a variável 'varre' da área de trabalho
end; % Fim do laço para cada ki(máx = 36)
%
% % % % Até aqui tem calculados todos os Ph(tau) de 1 grid de 36 medidas % % % %
%
I_Matriz_Alinhados = real(aux); % h(tau)real (Necessário para uso de WDEN / 1601 pontos)
Q_Matriz_Alinhados = imag(aux); % h(tau)imag (idem)
% Pegando os 801 primeiros pontos
for m = 1:No_Perfis;
    for n = 1:Nmeio;
        Pdb(m,n) = Potdb(m,n); % Ph(tau), em dB, até 1/2 perfil
        PPR(m,n) = PPRR(m,n); % |h(tau)|, relação de volts, até 1/2 perfil
    end;
end;
clear Potdb % Remove a variável 'Potdb' da área de trabalho
PPRoriginal = PPR; % |h(tau)|, até 1/2 perfil
% Normalizando os perfis pelo máximo
Max_PdB = max(Pdb,[],2); % Cria um vetor Max_PdB[No_Perfis] onde cada
% Elemento é o máximo de cada linha, no caso, máximo de cada perfil
for m = 1:No_Perfis
    Pdb(m,:) = Pdb(m,:)-Max_PdB(m); % Normaliza os perfis pelo máximo/dB
end
% Até aqui todos os perfis de retardo Ph(t,tau), em dB, normalizados
% pelo máximo, foram calculados para um mesmo local, em todos os grids(6ou7)
for ij = 1:No_Perfis;
    for ik = 1:Nmeio;
        Perfsujo(ij,ik) = Pdb(ij,ik); % 1/2 perfil, em dB, normalizado
    end
end
Perfsujo_completo = Perfsujo; % Para poder imprimir o perfil completo somente com o limiar
end; % Fim do laço para cada kp (máx = 6 ou 7, dependendo do local)
%
clear Pdb % Remove a variável 'Pdb' da área de trabalho
% Imprimindo os perfis
ent = 'a'; % Pergunta
while and(ent~='s',ent~='n')
    ent = input('Deseja visualizar alguns perfis medidos (sujos)? <s/n> ','s');
end
while ent == 's'

```

```

%clc; % Apaga a janela de comando para visualizar novo perfil
entnum = input('Digite o nº do perfil (entre 1 e 36): ');
if entnum <= No_Perfis;
    figure(entnum)
    plot(retardo,Perfsujo(entnum,:));ylabel...
    ('Potência relativa normalizada (dB)'),xlabel('T(nanosseg)'),...
    title(sprintf('Perfil Ruidoso N°%g Local%g',entnum,numlocal));
    zoom on
    grid minor
else
    disp('Número inválido. Intervalo válido [1:36]');
    disp(sprintf('Só se tem 36 perfis por grid, por LOCAL DE MEDIÇÃO'));
end
ent = 'a'; % Enquanto não teclar sim ou não, não continua
while and(ent~='s',ent~='n')
    ent = input('Deseja continuar a visualização dos perfis medidos (sujos)? <s/n> ','s');
end
end
%
% % % % % Imprimindo em 3D todos os perfis medidos em cada local (em 3D) % % % % %
Deltaeta = RetardoMax; % [nseg]
DurPerfis = (No_Perfis-1).*RetardoMax % Nanossegundos
tempo = 0:Deltaeta:DurPerfis; % nanossegundos
clear retardo % Remove a variável 'retardo' da área de trabalho
retardo = 0:deltaT:RetardoMax./2; % nseg/ 1/2 perfil = 801 pontos
%
% Imprimindo os perfis em 3D
ent = 'a'; % Pergunta
while and(ent~='s',ent~='n')
    ent = input('Deseja visualizar um local com todos os perfis medidos (sujos)? <s/n> ','s');
end
while ent == 's'
    %clc; % Apaga a janela de comando para visualizar novo perfil
    entnum = input('Digite o nº de perfis por local: ');
    if entnum <= No_Perfis
        figure(entnum)
        [retardo,tempo] = meshgrid(retardo,tempo);
        mesh(retardo,tempo,Perfsujo(:,1:Nmeio)),view(20,20),...
        xlabel('retardo-nanosseg'),ylabel('tempo-nanosseg'),zlabel('dB');
        zoom on
        grid minor
    else
        disp('Número inválido');
        % disp(sprintf('Existem apenas %g perfis por local.',No_Perfis));
        (sprintf('Existem apenas %g perfis por local.',No_Perfis));
    end
    ent = 'a'; % Enquanto não teclar sim ou não, não continua
    while and(ent~='s',ent~='n')
        ent = input('Deseja continuar essa visualização? <s/n> ','s');
    end
end
% % % % % % % % % % % SUAVIZAÇÃO DO RUÍDO % % % % % % % % % % % % % % % % %
%
% % % % % TRATAMENTO DE RUÍDO COM A TÉCNICA CFAR - Pág 32 % % % % %
% "Para efeito prático de ambiente indoor ->> 5% de multipercursos válidos"
% Suaviza o Perfil de Potência NORMALIZADO
disp(sprintf('Tratamento do ruído segundo Sousa! \n'))
% Suaviza o Perfil de Potência em dB, NORMALIZADO
% Segundo Sousa, para medidas outdoor: em geral 5% dos multipercursos são válidos
pmv = input('Qual é o percentual de multipercursos válidos (5% para ambiente Outdoor)? ');

```

```

if pmv == 0.05; % 5% de multipercursos válidos
    pmvdb = 10.*log10(-2.*log(0.5)); % = 1,4186// Desvio padrão do ruído térmico abaixo da mediana
else % Quando o percentual de multipercursos válidos é diferente de 5%
    pmvdb = 10.*log10(-2.*log(0.5.*(1-pmv))); % = 1,4811 se pmv = 0,01
end;
% eta = sqrt((-log(5e-6./2)./3));
% eta = 4.2997 = 6.3344 dB , = a Probabilidade de alarme falso,
% CFAR = 5e-6 = 0,0005% de chance de falso alarme, ou seja, 1 erro a cada 20000 amostras.
% Se CFAR = 5*10^-2, 1 erro a cada 20 amostras /etadb = 7,7753 dB
% Se CFAR = 10^-2, 1 erro a cada 100 amostras /etadb = 9,6428 dB
% Se CFAR = 5*10^-3, 1 erro a cada 200 amostras / etadb = 10,2517 dB
% Se CFAR = 5*10^-6, 1 erro a cada 200.000 amostras /etadb = 13,8761 dB
% Se CFAR = 5*10^-9, 1 erro a cada 200.000.000 amostras / etadb =15,8238 dB
eta = (-2.*log(5e-6)).^0.5; % 5e-6=Probabilidade de alarme falso, p/ o algoritmo usado
etadb = 20.*log10(eta) % (=20log(4,94)=13,8761 dB)
CFAR = input('Deseja alterar a Probabilidade de Alarme Falso? <s/n> ','s');
% eta = 4.9409
% etadb = 13.8761
%
if CFAR == 's'
    etadb = input('Qual o novo etadb desejado ( etadb padrão = 13.8761 - 27.2666 )? ');
end;
% Criando 1 vetor com as medianas dos perfis ruidosos
for m = 1:No_Perfis;
    Mediana_Perfil(m) = median(Perfsujo(m,:)); % Mediana em dBs
    Var_Ruido(m) = Mediana_Perfil(m)-pmvdb; % Variância do ruído
    Limiar_Ruido(m) = Var_Ruido(m)+etadb; % Limiar, em dB, abaixo do qual...
    % só se tem ruído. Tudo abaixo do limiar é posto ao nível do limite inferior.
end
% disp(sprintf('Limiar dos perfis do local %g',numlocal)), Limiar_Ruido(1:No_Perfis)% complexo
% Limite Inferior do perfil suavizado
% O limite inferior depende da relação adotada para Psinal/Pruído.
% Aqui é adotado nível da mediana como o limite inferior,
% Equivalendo a S/N>15 dB
Limite_Inferior = Mediana_Perfil;
% Passando o que está abaixo do limiar para o limite inferior
for m=1:No_Perfis;
    for n=1:Nmeio;
        if Perfsujo(m,n) <= Limiar_Ruido(m); % dB normalizado
            Perfsujo(m,n) = Limite_Inferior(m);
        end
    end
end
% Eliminando Ruído Impulsivo e tomando apenas amostras que satisfazem
% ao critério: um multipercurso válido (> limiar) deve se achar entre,
% pelo menos, 1 amostra maior que o limiar e ser > ambas.
for m=1:No_Perfis; % Para 1/2 Perfil
    amp(m,1)=Perfsujo(m,1);% dB normalizado
    amp(m,Nmeio)=Perfsujo(m,Nmeio);
    dnorm(m,1)=abs(Perfsujo(m,1)-Limite_Inferior(m));% desnormalizando
    dnorm(m,Nmeio)=0;
    jk=0;
    for n=2:Nmeio-1
        if( ( Perfsujo(m,n)> Perfsujo(m,n+1))&&( Perfsujo(m,n)>=...
            Perfsujo(m,n-1))&&( ( Perfsujo(m,n-1)>Limiar_Ruido(m))||...
            (Perfsujo(m,n+1)>Limiar_Ruido(m)) ) );
            amp(m,n)= Perfsujo(m,n);% Amplitudes dos Multipercursos Válidos
            dnorm(m,n)=abs( Perfsujo(m,n)-Limite_Inferior(m) ); % desnormalizando
            jk=jk+1; % No de Multipercursos Válidos
        else

```



```

        amp(m,n)=Limite_Inferior(m); % dB normalizado
        dnorm(m,n)=0;                % dBm desnormalizado
    end
end
                                % No de ecos válidos, por perfil m
disp(sprintf('No perfil %g temos %g multipercursos válidos.',m,jk))
end
for m=1:No_Perfis;                % Para 1/2 perfil
    for n=1:Nmeio;
        Limiar(m,n)=Limiar_Ruido(m); % dB, normalizado
        Limite(m,n)=Limite_Inferior(m);% dB, normalizado
    end
end
% Imprimindo os perfis (sujos só com Limiar de CFAR)
ent='a'; % Pergunta
while and(ent~='s',ent~='n')
    ent=input('Deseja visualizar alguns perfis medidos (sujos) com o limiar usando CFAR? <s/n> ','s');
end
while ent=='s'
    %clc;% Apaga a janela de comando para visualizar novo perfil
    entnum = input('Digite o nº do perfil: ');
    if entnum <= No_Perfis;
        figure(entnum)
        plot(retardo,Perfsujo_completo(entnum,:),retardo,...
            Limiar(entnum,:),'m');ylabel('Potência relativa normalizada (dB)',...
            xlabel('T(nanosseg)'),title(sprintf('Perfil Ruidoso N°%g Local%g',...
            entnum,numlocal));
        legend('\fontsize{7}Perfil medido','\fontsize{7}Limiar CFAR')
        zoom on
        grid minor
        %ylim([-100 5]); %%% Limite do eixo x
        %xlim([0 1100]); %%% Limite do eixo y
        hold on
    else
        disp('Número inválido');
        disp(sprintf('Só se tem 36 perfis por grid'));
    end
    ent='a'; % Enquanto não teclar sim ou não, não continua
    while and(ent~='s',ent~='n')
        ent=input('Deseja continuar a visualização dos perfis medidos (sujos) com o limiar usando CFAR?
<s/n> ','s');
    end
end
% Tirando multipercursos antes da LOS, levando eles para o nível do limite inferior
for m = 1:No_Perfis;                % Para 1/2 perfil
    imax(m) = find(amp(m,:) == max(amp(m,:)));% Retorna o índice onde a amp é max
    for i = 1:(imax(m)-1);
        amp(m,i) = Limite_Inferior(m);    % dnorm(m,i)=0;
    end
end
% Plotando os perfis que se deseja visualizar
clear entnum, clear ent
ent='a'; % Pergunta
while and(ent~='s',ent~='n')
    ent=input('Deseja visualizar alguns perfis sem ruído (CFAR)? <s/n> ','s');
end
while ent=='s';
    %clc;% Apaga a janela de comando para visualizar novo perfil
    entnum=input('Digite o nº do perfil: ');
    if entnum<=No_Perfis;

```

```

figure(entnum)
plot(retardo,amp(entnum,:), 'r',...
      retardo,Limiar(entnum), 'g-',...
      retardo,Limite(entnum,:), 'k-',...
      retardo,Perfsujo(entnum,:), 'g-',...
xlabel('\fontsize{12}Retardo (ns)');ylabel('\fontsize{12}Potência Relativa (dB)');
titulo = (sprintf('Perfis Sujo e Limpo N°%g Local%g',entnum,numlocal));title(titulo)
legend('\fontsize{10}Perfil Sujo medido','\fontsize{10}Limiar CFAR',...
       '\fontsize{10}Pontos válidos','\fontsize{10}Perfil Limpo medido')
zoom on
grid minor
%set(gcf,'color','w'),set(gca,'fontsize',8),zoom on %, grid on
else
disp('N° inválido');
end
ent='a'; % Enquanto não teclar sim ou não, não continua
while and(ent~='s',ent~='n')
ent=input('Deseja continuar essa visualização? <s,n> ','s');
end
end
% Imprimindo todos os perfis limpos de cada local, usando CFAR (em 3D)
clear retardo tempo % Remove a variável 'retardo' da área de trabalho
retardo = 0:deltaT:RetardoMax./2; % nseg
tempo=0:Deltaeta:DurPerfis; % microssegundos
% Imprimir perfis limpos em 3D
ent='a'; % Pergunta
while and(ent~='s',ent~='n')
ent=input('Deseja visualizar um local com todos os perfis limpos (CFAR)? <s/n> ','s');
end
while ent=='s'
%clc;% Apaga a janela de comando para visualizar novo perfil
entnum=input('Digite o nº do local: ');
if entnum<=numlocal;
figure(No_Perfis+entnum)
[retardo,tempo]=meshgrid(retardo,tempo);
mesh(retardo,tempo,dnorm(:,1:Nmeio)),...%
mesh(retardo,tempo,dnorm(:,1:Nmeio),retardo,Limiar(entnum,:)),...
view(20,20),xlabel('nanoseg'),ylabel('tempo-nanosseg'),zlabel('dB')
zoom on
grid minor
else
disp('Número inválido');
disp(sprintf('Existem apenas %g locais.',numlocal));
end
ent='a'; % Enquanto não teclar sim ou não, não continua
while and(ent~='s',ent~='n')
ent=input('Deseja continuar essa visualização? <s/n> ','s');
end
end
% Transformando os perfis limpos para relação de WATTS
for m = 1:No_Perfis
for n = 1:Nmeio
if dnorm(m,n)~=0
Perfis_Sem_Ruido_SOUSAW(m,n)=(10.^(dnorm(m,n)/10)); % Perfis sem ruído, relação de W
else
Perfis_Sem_Ruido_SOUSAW(m,n)=0; % Como não aparecem no somatório, podem ser feitos nulos
end;
end;
end;
% Cálculo dos parâmetros para comparação (retardo médio e espalhamento de retardo) para CFAR

```

```

for m=1:No_Perfis,i=1;
  for n=1:Nmeio
    if Perfis_Sem_Ruido_SOUSA(m,n)~=0
      contador(m,i)=n;
      PSemRuidoW_SOUSA(m,n)= Perfis_Sem_Ruido_SOUSA(m,n); %
      i=i+1;
    else
      PSemRuidoW_SOUSA(m,n)=0; % Antes estava 0 WATTS
    end;
  end;
end;
%
% % % % % % % % % % CÁLCULO DO RETARDO MÉDIO % % % % % % % %
for m = 1:No_Perfis
  Denominador(m) = 0;
end
for m = 1:No_Perfis
  Numerador = 0;
  for n = 1:Nmeio
    if PSemRuidoW_SOUSA(m,n)~=0; % Qdo o perfil é nulo, nada soma aos termos
      Numerador = Numerador + (n-contador(m,1)).*...
      deltaT.*PSemRuidoW_SOUSA(m,n);% Perfil Limpo em Watts
      Denominador(m) = Denominador(m) + PSemRuidoW_SOUSA(m,n);
    end;
  end;
  Retardo_Medio_SOUSA(m) = Numerador./Denominador(m);%nseg
end;
disp(sprintf('Média dos Retardos Médios')),medret=mean(Retardo_Medio_SOUSA)
%
% % % % % % % % % % CÁLCULO DO ESPALHAMENTO DE RETARDO/PERFIL % % % % % % % %
for m = 1:No_Perfis; % 36 Perfis
  Numerador = 0; % O denominador é o mesmo calculado anteriormente
  for n = 1:Nmeio
    if PSemRuidoW_SOUSA(m,n)~=0; % Qdo o perfil é nulo, nada soma aos termos
      Numerador=Numerador+(((n-contador(m,1)).*...
      deltaT-Retardo_Medio_SOUSA(m)).^2).*PSemRuidoW_SOUSA(m,n);
    end;
  end
  SigmaT_SOUSA(m) = sqrt(Numerador./Denominador(m)); % nanoseg
end
disp(sprintf('Média dos Espalhamentos de Retardos')), medesp=mean(SigmaT_SOUSA)
clear Numerador Denominador % Remove as variáveis 'Numerador' e 'Denominador' da área de trabalho
disp(sprintf('Retardo médio dos perfis do local %g',numlocal)), Retardo_Medio_SOUSA
disp(sprintf('Espalhamento de retardo dos perfis do local %g', numlocal)),SigmaT_SOUSA
Desvio = std(SigmaT_SOUSA)
%
% % % % % % % % % % FIM % % % % % % % % % % % % % % % % % %

```

Anexo B – Programa em Matlab: criação da matriz base de características

Segue o código do programa em Matlab usado criação da matriz base de características.

Nota: para caracterização do CRM com uso do HMM, é necessário ter o HMM *Toolbox* para Matlab.

```
% Universidade Federal Fluminense (UFF)
% Pós-Graduação em Engenharia Elétrica e de Telecomunicações (PPGEET)
% Tese: Aplicação de HMM na modelagem de CRM - Leonardo Ribeiro (2019)
%
clear % Remove variáveis da área de trabalho. Libera memória do sistema.
close all % Exclui todas as figuras cujas alças não estejam ocultas
clc; % Limpa a janela de comando (Clear Command Window)
%
% Parâmetros da campanha de Medição/Sondagem
% Técnica de Sondagem no Domínio da Frequência varreu a banda de 960 ~ 1710MHz
% ou 750MHz em 1601 amostras de frequência em pontos discretos.
% "A medição desta banda em vários pontos do ambiente permite determinar a função T(f,t) do canal,
%
% Setup De Medição
B = 750; % [MHz] - Banda de frequência sondada = 1710 - 960 MHz
pontos = 1601; % Varredura de B com 1601 amostras de frequência em cada ponto
Nmeio = (pontos+1)/2; % 801 pontos/amostras de medição
deltaF = B/(pontos-1); % [MHz] Resolução na frequência (intervalo entre as amostras) = 0,46875MHz
%
RetardoMax = 10^3/deltaF; % [nseg] -> Duração de 1 perfil = 2,1333 x 10^3 nseg
% RetardoMax = tempo máx entre 2 medições em um mesmo ponto
% deltaF em [MHz]. RetardoMax em [micros] = 10^-6
% RetardoMax em [nseg] = 10^-9 -> RetardoMax = (1x10^3)/(deltaFx10^3)
%
deltaT = 10^3/B; % [nseg] -> Resolução temporal = 1,3333
% Mesmo raciocínio. x10^3 para alterar a unidade para [nseg]
% [Christiano 2006 - Pág 65] deltaT indica a diferença mínima no tempo de
% chegada entre 2 raios, p/ que possam ser distinguidos pelo analisador.
% Ou seja, representa a distância mínima percorrida por dois raios, para a
% qual o analisador ainda consegue diferenciá-lo.
%
% Tperfis=deltaT*(Nmeio-1); %duracao de 1/2 perfil de retardos
% cont=1;
% totperfis=0;
%
% % % % % % CHAMANDO OS ARQUIVOS GRAVADOS DA MEDIÇÃO % % % % % %
%
% numlocal = input('Entre com o nº do LOCAL DE MEDIÇÃO desejado (RX que pode variar entre 1 e 7): ');
input.state{1,1} = zeros(36,2);
input.state{1,2} = zeros(36,2);
input.state{1,3} = zeros(36,2);
input.state{1,4} = zeros(36,2);
%
for rx = 1:4
    %
    % Nº do LOCAL DE MEDIÇÃO representa cada um dos 7 pontos de medição (RX1 ~ RX7)
```

```

% A medição em cada LOCAL DE MEDIÇÃO foi feita com GRIDS de 36 posições.
% 36 funções (1 função para ponto) foram obtidas para cada GRID
% (cada GRID é uma matriz 6x6 com 15cm de afastamento entre os pontos)
% Freq banda superior = 1710MHz >> lambda = 17,54cm >> 0,38lambda = 6,7cm
% Freq banda inferior = 960MHz >> lambda = 31,25cm >> 0,38lambda = 11,9cm
% 1a correlação zero entre sinais (sinais descorrelacionados) = 0,38 lambda
% Na prática, utiliza-se 0,50 lambda (aqui usou 0,38 lambda)
% Esse método possibilita a obtenção de estatísticas do canal, já que em
% cada ponto do grid a RX do sinal está descorrelacionada dos demais
%
% Loop (kp) para incrementar até o nº máximo de GRIDS num local(numlocal)
%
Amp = zeros([36 1601]); % INSERI MATRIZ DE ZEROS
T = zeros([36 1601]); % INSERI MATRIZ DE ZEROS
Vetor = zeros([1 1601]); % INSERI MATRIZ DE ZEROS
aux = zeros([36 1601]); % INSERI MATRIZ DE ZEROS
%
No_Perfis = 36; % 36 pontos de medição em cada LOCAL DE MEDIÇÃO
%
numlocal = rx;
%
for kp = numlocal; % kp = Nº do LOCAL DE MEDIÇÃO
    %No_Perfis=36; % Grid 36 pnts de medição em cada LOCAL DE MEDIÇÃO
    pp=int2str(kp); % Conversão p/ string p/ criar nome 'ponto1_1.mat'
    nome=['ponto',pp];% Variável 'nome' recebe string ponto + pp
    % Se pp = 1 -> nome = 'ponto1'
    % Loop para incrementar até No_Perfis medidos em cada GRID (Ki)
    %
    for ki = 1:No_Perfis; % ki é o nº do perfil (= ponto do grid)
        rr = int2str(ki); % Conversão para string
        arquivo = [nome,'_',rr];% Criação do nome do arquivo 'ponto1_1'
        load(arquivo); % Carrega 'arquivo' correspondente para o MatLab
        % Exemplo: para kp = 1, então carrega 'arquivo' com nomes 'ponto1_1' até 'ponto1_36'
        % Cada arquivo 'ponto1_1' até 'ponto1_36' é uma matriz M de 2x1601
        % 1a linha corresponde as amplitudes RX dos sinais (M 1x1601)
        % 2a linha corresponde as fases RX dos sinais (M 2x1601)
        Amp2(1,:) = varre(1,:); % [dB]. Gera array com 1601 medições
        Fase2(1,:) = varre(2,:); % Em graus. 2 Linha array com 1601 medições
        % Varre(1,:) = Retorna a 1a linha, Varre(2,:) = Retorna a 2a linha
        % Carrega no MATLAB as medições escolhidas com 'load(arquivo)'.
        % Valores armazenados nas variáveis Amp2 e Fase2.
        % Amp2 = Array com medições de potência em dB.
        % Fase2 = Array com medições de Fase em graus.
        % Amp2 e Fase2 geram uma matriz 2x1601 com a variável Varre
        %
        % ponto11 = varre % para visualização <<-----AQUI
        % clear varre; % Remove a variável 'varre' da área de trabalho,
        % libera memória do sistema. 'varre' armazena Amp2 e Fase2
        %
        % Parâmetros Estatísticos da Sondagem
        %
        retardo = 0:deltaT:RetardoMax./2; % nseg/ 1/2 perfil = 801 pontos
        % x:y:z -> Início em x. Fim em z. De x até z com incremento de y.
        % De 0 até [(2,1333x10^3 nseg)/2] com incremento de 1,3333
        %
        % Composição dos dados em números complexos
        %
        Amp(ki,:) = 10.^(Amp2(1,+)/20); % [V] Amplitude em volts,
        % pois medição(Amp2) está em dB
        % Amp = Matrix 36x1601. Para o ponto de medição RX1, a variável Amp

```

```

% está mostrando as 1601 medições de amplitude [V] de cada ponto do
% Grid. 1601 medições em cada um dos 36 pontos do Grid.
%
T(ki,:) = Amp(ki,:).*cos(Fase2(1,:)*pi/180)+i*Amp(ki,:).*...
sin(Fase2(1,:)*pi/180); % T(f,t) -> Pág 36
% pi/180 = fase em radianos
% Pois MATLAB calcula o COS e SEN de rad e não de graus
%
% Relação de VOLTS x FREQ
%
% Transformada De Fourier de um sinal finito da margem para
% espúrios (parte ã real) no domínio da frequência (conhecido
% como Vazamento de Frequência Espectral - Spectral Leakage).
%
% Aplicando a Janela de Blackman Harris para reduzir espúrios da
% resposta do canal no domínio dos retardos (associados a
% intervalos finitos de observação no domínio da frequência),
% antes da Transformada de Fourier os dados discretos são aplicados
% a uma função janela para redução dos lóbulos secundários.
% Janela de Blackman Harris possui uma das melhores performances.
%
a0 = 0.42323; % Coeficientes 1 Janela Blackman Harris c/ 3 termos
a1 = 0.49755; % Coeficientes 2 Janela Blackman Harris c/ 3 termos
a2 = 0.07922; % Coeficientes 3 Janela Blackman Harris c/ 3 termos
% Vetor = blackmanharris(pontos,'periodic'); % Array/matriz coluna
% wvtool(Vetor)
% PBH = Vetor'; % Gera array/matriz linha a partir do array Vetor
%
% Função blackmanharris() % Minimum 4-term Blackman-Harris window
% Janela Blackman Harris c/ 4 termos (a0, a1, a2 e a3)
% a0 = 0.35875, a1 = 0.48829, a2 = 0.14128, a3 = 0.01168
%
for n = 0:(pontos-1)
    Vetor(n+1) = a0 - a1*cos(2*pi*n/pontos) + a2*cos(2*pi*2*n/pontos);
end
PBH = Vetor; % Gera array com 1601 medições armazenadas em Vetor
clear Vetor % Remove a variável 'Vetor' da área de trabalho, libera memória do sistema
%
% % % % % % % % APLICANDO A IFFT A T(f,t) P/ CHEGAR A h(t,tau) % % % % % % %
aux(ki,:) = ifft( (T(ki,:)).*PBH,pontos ); % h(tau)
% h(tau) - relação de volts x retardo -> 1601 pontos
PPRR = abs(aux); % |h(tau)|/1601 pontos
Pot = PPRR.^2; % Ph(tau)/Relação de WATTS/ 1601 pontos
Potdb = 10.*log10(Pot); % Perfil de Potencia relativo Ph(tau), em dB/ 1601 pontos
clear Pot % Remove a variável 'varre' da área de trabalho
end; % Fim do laço para cada ki(máx = 36)
%
% % % % Até aqui tem calculados todos os Ph(tau) de 1 grid de 36 medidas % % % %
%
I_Matriz_Alinhados=real(aux); % h(tau)real (Necessário para uso de WDEN / 1601 pontos)
Q_Matriz_Alinhados=imag(aux); % h(tau)imag (idem)
% Pegando os 801 primeiros pontos
for m=1:No_Perfis;
    for n=1:Nmeio;
        Pdb(m,n)=Potdb(m,n); % Ph(tau), em dB, até 1/2 perfil
        PPR(m,n)=PPRR(m,n); % |h(tau)|, relação de volts, até 1/2 perfil
    end;
end;
clear Potdb % Remove a variável 'Potdb' da área de trabalho
PPRoriginal=PPR; % |h(tau)|, até 1/2 perfil

```

```

% Normalizando os perfis pelo máximo
Max_PdB = max(Pdb,[],2); % Cria um vetor Max_PdB[No_Perfis] onde cada
% Elemento é o máximo de cada linha, no caso, máximo de cada perfil
for m=1:No_Perfis
    Pdb(m,:)=Pdb(m,)-Max_PdB(m); % Normaliza os perfis pelo máximo/dB
end
% Até aqui todos os perfis de retardo Ph(t,tau), em dB, normalizados
% pelo máximo, foram calculados para um mesmo local, em todos os grids(6ou7)
for ij=1:No_Perfis;
    for ik=1:Nmeio;
        Perfsujo(ij,ik)=Pdb(ij,ik);% 1/2 perfil, em dB, normalizado
    end
end
Perfsujo_completo=Perfsujo; % Para poder imprimir o perfil completo somente com o limiar
end; % Fim do laço para cada kp (máx = 6 ou 7, dependendo do local)
clear Pdb % Remove a variável 'Pdb' da área de trabalho
% Imprimindo os perfis
%
ent='n'; % Pergunta
% ent='a'; % Pergunta
% while and(ent~='s',ent~='n')
%     ent=input('Deseja visualizar alguns perfis medidos (sujos)? <s/n>','s');
% end
while ent=='s'
    %clc; % Apaga a janela de comando para visualizar novo perfil
    entnum=input('Digite o nº do perfil: ');
    if entnum <= No_Perfis;
        figure(entnum)
        plot(retardo,Perfsujo(entnum,:));ylabel...
        ('Potência relativa normalizada (dB)'),xlabel('T(nanosseg)'),...
        title(sprintf('Perfil Ruidoso N°%g Local%g',entnum,numlocal));
        zoom on
        grid minor
        %ylim([-100 5]); % Limite do eixo x % Alteração minha
        %xlim([0 1100]); % Limite do eixo y % Alteração minha
        %set(gca,'xtick',0:deltaT:RetardoMax/2) NÃO USEI. PRECISO REFINAR.
        %set(gca,'ytick',[0:50:100]) NÃO USEI
    else
        disp('Número inválido. Intervalo válido [1:36]');
        disp(sprintf('Só se tem 36 perfis por grid, por LOCAL DE MEDIÇÃO'));
    end
    ent = 'n'
%     ent='a';% Enquanto não teclar sim ou não, não continua
%     while and(ent~='s',ent~='n')
%         ent=input('Deseja continuar a visualização dos perfis medidos (sujos)? <s/n>','s');
%     end
end
% % % % % Imprimindo em 3D todos os perfis medidos em cada local (em 3D) % % % % %
Deltaeta = RetardoMax; % nseg
DurPerfis = (No_Perfis-1).*RetardoMax % nseg
tempo = 0:Deltaeta:DurPerfis; % nseg
clear retardo % Remove a variável 'retardo' da área de trabalho
retardo = 0:deltaT:RetardoMax./2; % nseg/ 1/2 perfil = 801 pontos
%
% Imprimindo os perfis em 3D
ent='n'; % Pergunta
% ent='a'; % Pergunta
% while and(ent~='s',ent~='n')
%     ent=input('Deseja visualizar um local com todos os perfis medidos (sujos)? <s/n>','s');
% end

```



```

        Perfsujo(m,n)=Limite_Inferior(m);
    end
end
end
% Eliminando Ruído Impulsivo e tomando apenas amostras que satisfazem
% ao critério:um multipercurso válido(> limiar) deve se achar entre,
% pelo menos, 1 amostra maior que o limiar e ser > ambas.
for m=1:No_Perfis;      % Para 1/2 Perfil
    amp(m,1) = Perfsujo(m,1);% dB normalizado
    amp(m,Nmeio) = Perfsujo(m,Nmeio);
    dnorm(m,1) = abs(Perfsujo(m,1)-Limite_Inferior(m));% desnormalizando
    dnorm(m,Nmeio) = 0;
    jk = 0;
    for n=2:Nmeio-1
        if( ( Perfsujo(m,n)> Perfsujo(m,n+1))&&( Perfsujo(m,n)>=...
            Perfsujo(m,n-1))&&( ( Perfsujo(m,n-1)>Limiar_Ruido(m))||...
            (Perfsujo(m,n+1)>Limiar_Ruido(m)) ) );
            amp(m,n) = Perfsujo(m,n);% Amplitudes dos Multipercursos Válidos
            dnorm(m,n) = abs( Perfsujo(m,n)-Limite_Inferior(m) );% desnormalizando
            jk = jk + 1; % No de Multipercursos Válidos
        else
            amp(m,n) = Limite_Inferior(m);% dB normalizado
            dnorm(m,n) = 0;          % dBm desnormalizado
        end
    end
end
% No de ecos válidos, por perfil m
disp(sprintf('No perfil %g temos %g multipercursos válidos.',m,jk))
end
for m = 1:No_Perfis;      % Para 1/2 perfil
    for n = 1:Nmeio;
        Limiar(m,n) = Limiar_Ruido(m); % dB, normalizado
        Limite(m,n) = Limite_Inferior(m);% dB, normalizado
    end
end
% Imprimindo os perfis (sujos só com Limiar de CFAR)
ent = 'n'
% ent='a'; % Pergunta
% while and(ent~='s',ent~='n')
%     ent=input('Deseja visualizar alguns perfis medidos(sujos)com o limiar usando CFAR? <s/n>','s');
% end
while ent=='s'
    %clc; % Apaga a janela de comando para visualizar novo perfil
    entnum=input('Digite o nº do perfil: ');
    if entnum <= No_Perfis;
        figure(entnum)
        plot(retardo,Perfsujo_completo(entnum,:),retardo,...
            Limiar(entnum,:),'m');ylabel('Potência relativa normalizada (dB)'),...
            xlabel('T(nanosseg)'),title(sprintf('Perfil Ruidoso N°%g Local%g',...
            entnum,numlocal));
        legend('\fontsize{7}Perfil medido',\fontsize{7}Limiar CFAR')
        zoom on
        grid minor
        % ylim([-100 5]); %%% Limite do eixo x
        % xlim([0 1100]); %%% Limite do eixo y
        hold on
    else
        disp('Número inválido');
        disp(sprintf('Só se tem 36 perfis por grid'));
    end
end
ent = 'a';% Enquanto não teclar sim ou não, não continua

```

```

while and(ent~='s',ent~='n')
    ent=input('Deseja continuar a visualização dos perfis medidos (sujeitos) com o limiar usando CFAR?
<s/n> ','s');
end
end
% Tirando multipercursos antes da LOS, levando eles p/o nível do limite inferior
for m=1:No_Perfis% Para 1/2 perfil
    imax(m) = find(amp(m,:) == max(amp(m,:)));% Retorna o índice onde a amp é max
    for i = 1:(imax(m)-1)
        amp(m,i)=Limite_Inferior(m);% dnorm(m,i)=0;
    end
end
end
%Plotando os perfis que se deseja visualizar
clear entnum
clear ent
ent = 'n'
% ent='a'; % Pergunta
% while and(ent~='s',ent~='n')
%     ent=input('Deseja visualizar alguns perfis sem ruído(CFAR)? <s/n> ','s');
% end
while ent=='s';
    %clc; % Apaga a janela de comando para visualizar novo perfil
    entnum=input('Digite o nº do perfil: ');
    if entnum<=No_Perfis;
        figure(entnum)
        plot(retardo,amp(entnum,:),'r-',retardo,Limiar(entnum),'g-!...',
            retardo,Limite(entnum,:),'k-',retardo,...
            Perfsujo(entnum,:),'g-'),...
            xlabel('\fontsize{12}Retardo (ns)');ylabel('\fontsize{12}Potência Relativa (dB)');
        titulo = [sprintf('Perfis Sujo e Limpo Nº%g Local%g',entnum,numlocal)];title(titulo)
        legend('\fontsize{10}Perfil Sujo medido','\fontsize{10}Limiar CFAR',...
            '\fontsize{10}Pontos válidos','\fontsize{10}Perfil Limpo medido')
        % set(gcf,'color','w'),set(gca,'fontsize',8),zoom on %, grid on
    else
        disp('Nº inválido');
    end
    ent='a';% Enquanto não teclar sim ou não, não continua
    while and(ent~='s',ent~='n')
        ent=input('Deseja continuar essa visualização?<s,n> ','s');
    end
end
end
% Imprimindo todos os perfis limpos de cada local, usando CFAR(em 3D)
clear retardo tempo % Remove a variável 'retardo' da área de trabalho
retardo = 0:deltaT:RetardoMax./2;% nseg
tempo = 0:Deltaeta:DurPerfis; % microssegundos
%Imprimir perfis limpos em 3D
ent = 'n'
% ent='a'; % Pergunta
% while and(ent~='s',ent~='n')
%     ent=input('Deseja visualizar um local com todos os perfis limpos(CFAR)? <s/n> ','s');
% end
while ent=='s'
    %clc; %Apaga a janela de comando para visualizar novo perfil
    entnum=input('Digite o nº do local: ');
    if entnum<=numlocal;
        figure(No_Perfis+entnum)
        [retardo,tempo]=meshgrid(retardo,tempo);
        mesh(retardo,tempo,dnorm(:,1:Nmeio)),...%
        mesh(retardo,tempo,dnorm(:,1:Nmeio),retardo,Limiar(entnum,:)),...
        view(20,20),xlabel('nanoseg'),ylabel('tempo-nanosseg'),zlabel('dB')
    end
end

```

```

zoom on
grid minor
else
    disp('Número inválido');
    (sprintf('Existem apenas %g locais.',numlocal));
end
ent='a'; % Enquanto não teclar sim ou não, não continua
while and(ent~='s',ent~='n')
    ent=input('Deseja continuar essa visualização? <s/n> ','s');
end
end
% Transformando os perfis limpos para relação de WATTS
for m=1:No_Perfis
    for n=1:Nmeio
        if dnorm(m,n)~=0
            Perfis_Sem_Ruido_SOUSA(m,n)=(10.^(dnorm(m,n)/10)); %Perfis sem ruído, relação de W
        else
            Perfis_Sem_Ruido_SOUSA(m,n)=0; % Como estes não aparecem no somatório, podem ser
feitos nulos
        end;
    end;
end;
% Cálculo dos parâmetros para comparação (retardo médio e espalhamento de retardo)para CFAR
for m=1:No_Perfis,i=1;
    for n=1:Nmeio
        if Perfis_Sem_Ruido_SOUSA(m,n)~= 0
            contador(m,i) = n;
            PSemRuidoW_SOUSA(m,n) = Perfis_Sem_Ruido_SOUSA(m,n); %
            i=i+1;
        else
            PSemRuidoW_SOUSA(m,n) = 0;% Antes estava 0 WATTS
        end;
    end;
end;
% % % % % % % % % % % % % % CÁLCULO DO RETARDO MÉDIO % % % % % % % % % % % %
for m = 1:No_Perfis
    Denominador(m)=0;
end
for m=1:No_Perfis
    Numerador = 0;
    for n=1:Nmeio
        if PSemRuidoW_SOUSA(m,n)~= 0;% Qdo o perfil é nulo, nada soma aos termos
            Numerador = Numerador + (n-contador(m,1)).*...
                deltaT.*PSemRuidoW_SOUSA(m,n);% Perfil Limpo em Watts
            Denominador(m) = Denominador(m) + PSemRuidoW_SOUSA(m,n);
        end
    end
    Retardo_Medio_SOUSA(m) = Numerador./Denominador(m);% nseg
end;
disp(sprintf('Média dos Retardos Médios')),medret = mean(Retardo_Medio_SOUSA)
%
% % % % % % % % % % % % % % CÁLCULO DO ESPALHAMENTO DE RETARDO/PERFIL % % % % % % % % % %
for m=1:No_Perfis; % 36 Perfis
    Numerador = 0; % O denominador é o mesmo calculado anteriormente
    for n=1:Nmeio
        if PSemRuidoW_SOUSA(m,n)~=0; % Qdo o perfil é nulo, nada soma aos termos
            Numerador=Numerador+(((n-contador(m,1)).*...
                deltaT-Retardo_Medio_SOUSA(m)).^2).*PSemRuidoW_SOUSA(m,n);
        end;
    end
end
end

```

```

SigmaT_SOUSA(m) =sqrt(Numerador./Denominador(m));% nanoseg
end
disp(sprintf('Média dos Espalhamentos de Retardos')), medesp=mean(SigmaT_SOUSA)
clear Numerador Denominador% Remove as variáveis 'Numerador' e 'Denominador' da área de trabalho
disp(sprintf('Retardo médio dos perfis do local %g',numlocal)), Retardo_Medio_SOUSA
disp(sprintf('Espalhamento de retardo dos perfis do local %g', numlocal)),SigmaT_SOUSA
Desvio = std(SigmaT_SOUSA)
%
% InitDataSet
% Não é necessário criar a Matriz de Características da Região RX dinamicamente.
if rx == 1
    input.state{ 1,1 }(:,1) = Retardo_Medio_SOUSA';
    input.state{ 1,1 }(:,2) = SigmaT_SOUSA';
%
elseif rx == 2
    input.state{ 1,2 }(:,1) = Retardo_Medio_SOUSA';
    input.state{ 1,2 }(:,2) = SigmaT_SOUSA';
%
elseif rx == 3
    input.state{ 1,3 }(:,1) = Retardo_Medio_SOUSA';
    input.state{ 1,3 }(:,2) = SigmaT_SOUSA';
%
elseif rx == 4
    input.state{ 1,4 }(:,1) = Retardo_Medio_SOUSA';
    input.state{ 1,4 }(:,2) = SigmaT_SOUSA';
%
end
%
end
%
save('InitDataSet_A7.mat','input');
%
%%%%%%%%%%%%%% FIM %%%%%%%%%%%%%%%

```

Anexo C – Programa em Matlab: Processamento dos dados com HMM *Toolbox*

O programa abaixo analisa o conjunto de transições como um todo para cada ambiente de medição. O código principal é `DataProcessing.m` que representa a implementação da rotina Hidden Markov Model (HMM) *Toolbox* para Matlab. `DataProcessing.m` chama outras rotinas como `hmmBuilder.m`, que gera a matriz de probabilidade de estados, e `fillResults.m`, que calcula determinados parâmetros estatísticos usados como base para a verificação dos modelos. Por sua vez, `hmmBuilder.m` e `fillResults.m` também chamam outras rotinas do HMM *Toolbox* para Matlab.

- `DataProcessing.m`

```
% Universidade Federal Fluminense (UFF)
% Pós-Graduação em Engenharia Elétrica e de Telecomunicações (PPGEET)
%
% Código para Topologia de Bakis & Processamento dos dados HMM Toolbox
% Gerar uma variação aleatória entre os estados 1 e 7
% Topologia de Bakis permite transição para estado subsequente ou salto de apenas um estado.
% Ela apresenta uma estrutura temporal e restringe as transições entre estados,
% Traz maior eficiência para as iterações e treinamentos necessários e mantendo ainda certa flexibilidade.
%
clear;    % Remove variáveis da área de trabalho. Libera memória do sistema.
clc;     % Limpa a janela de comando (Clear Command Window)
close all; % Exclui todas as figuras cujas alças não estejam ocultas
%
%% Definição do que se deseja estudar/observar
prompt = input(' 1 = Topologia Bakis aleatória.\n 2 = Topologia Bakis pré-definida.\n');
NetType = input(' Entre com nº desejado: ');
%
if NetType == 1 % Escolha 'Topologia Bakis aleatória' pelo usuário
    %
    indexLenght = 1:1:3;
    k = 1;
    s = 1;
    ArrayIndex(1,s) = k;
    s = s+1;
    while k < 3
        randomGridStateIndex = randperm(indexLenght(3)); % Gera um número aleatório entre 1, 2 e 3
        randomGridStateIndex = randomGridStateIndex(1);
        while randomGridStateIndex < k
            randomGridStateIndex = randperm(indexLenght(3));
            randomGridStateIndex = randomGridStateIndex(1);
        end
        ArrayIndex(1,s) = randomGridStateIndex;
        k = randomGridStateIndex;
        s = s+1;
    end
    %
    clear randomGridStateIndex;
```

```

indexLenght1 = 3:1:5;
while k < 5
    randomGridStateIndex1 = randperm(indexLenght1(3)); % Gera um número aleatório entre 3, 4 e 5
    randomGridStateIndex1 = randomGridStateIndex1(1);
    while randomGridStateIndex1 < k
        randomGridStateIndex1 = randperm(indexLenght1(3));
        randomGridStateIndex1 = randomGridStateIndex1(1);
    end
    ArrayIndex(1,s) = randomGridStateIndex1;
    k = randomGridStateIndex1;
    s = s+1;
end
%
clear randomGridStateIndex1;
indexLenght2 = 5:1:7;
while k < 7
    randomGridStateIndex2 = randperm(indexLenght2(3)); % Gera um número aleatório entre 5, 6 e 7
    randomGridStateIndex2 = randomGridStateIndex2(1);
    while randomGridStateIndex2 < k
        randomGridStateIndex2 = randperm(indexLenght2(3));
        randomGridStateIndex2 = randomGridStateIndex2(1);
    end
    ArrayIndex(1,s) = randomGridStateIndex2;
    k = randomGridStateIndex2;
    s = s+1;
end
%
clear randomGridStateIndex2;
indexLenght3 = 7:1:8;
while k <= 7
    randomGridStateIndex3 = randperm(indexLenght3(2)); % Gera um número aleatório entre 7 e 8
    randomGridStateIndex3 = randomGridStateIndex3(1);
    while randomGridStateIndex3 < k
        randomGridStateIndex3 = randperm(indexLenght3(2));
        randomGridStateIndex3 = randomGridStateIndex3(1);
    end
    ArrayIndex(1,s) = randomGridStateIndex3; % Array com valores aleatórios de 1 até 8
    k = randomGridStateIndex3;
    s = s+1;
end
% Deve-se retirar o valor 8. Deve-se ir de 1 até 7 (Região 1 até Região 7)
j = 1;
while ArrayIndex(1,j)<8
    ArrayBakisTopology(1,j) = ArrayIndex(1,j); % Array com valores aleatórios para Topologia de Bakis
    j = j+1;
end
clear randomGridStateIndex3 indexLenght indexLenght1 indexLenght2 indexLenght3
%
% Array com Topologia Bakis gerada e salva em ArrayBakisTopology
%
% FIM DO PROGRAMA QUE CRIA TOPOLOGIA BAKIS
%
end
%
if NetType == 2 % Escolha 'Topologia Bakis pré-definida' pelo usuário
    ArrayBakisTopology = [1 2 2 3 4 4 5 6 6 7];
end
%
k2 = 1; % Variável para incremento das funções abaixo
LengthBakisTopology = length(ArrayBakisTopology); % Comprimento Topologia Bakis

```

```

%
MatrixNameBakisZeros = zeros(2,2,36);
ArrayBakis_RegionZeros = zeros(36,2);
%
ArrayBakis_RegionMatrix = zeros(LengthBakisTopology - 1,12);
ArrayBakis_Region = zeros(LengthBakisTopology,3);
%
for k=1:LengthBakisTopology
    if k <= LengthBakisTopology - 1          % '-1' devido passagem de Rx atual para Rx seguinte
        x = int2str(ArrayBakisTopology(k));   % Rx atual
        x2 = int2str(ArrayBakisTopology(k+1)); % Rx seguinte
        name = ['Rx',x,'ToRx',x2,'data'];     % Cria str 'Rx?ToRx?data'
        ArrayBakis_RegionMatrix(k,:) = name;  % Cria str Rx? e abaixo var Rx?
        assignin('base',char(name),MatrixNameBakisZeros)
        %
        % Abaixo: contagem das Regiões sem repetição.
        % Apenas para saber quais regiões participaram da contagem.
        Array1 = ArrayBakisTopology(k);       % Apenas para criar variável 'Array' sem repetição
        Array2 = ArrayBakisTopology(k+1);     % Apenas para criar variável 'Array' sem repetição
        if Array1 ~= Array2                   % 'Array' conta cada 'Rx' de 1 a 7 sem repetição
            Array(k2) = ArrayBakisTopology(k); % Array com índice k2
            Array(k2+1) = ArrayBakisTopology(k+1); % Array com índice k2+1
            k2 = k2+1;
        end
    end
    %
end
%
if k <= max(size(ArrayBakisTopology))        % size(ArrayBakisTopology)=1 10. 1 linha 10 colunas.
    x = int2str(ArrayBakisTopology(k));
    ArrayBakis_Region(k,:) = ['Rx',x];        % Cria str Rx? e abaixo var Rx?
    assignin('base',char(ArrayBakis_Region(k,:)),ArrayBakis_RegionZeros)
end
%
end
%
clear x x2 k1 k2          % Limpando memória. Variáveis não são mais necessárias.
clear Array1 Array2 name % Limpando memória. Variáveis não são mais necessárias.
%
% Até aqui, defini as variáveis necessárias e atribui matriz de zeros.
%
input = load ('InitDataSet');
% InitDataSet.mat = Tabela com Retardos Médios e Espalhamentos de Retardo
% de cada um dos pontos de medição do sinal: RX1, RX2, ... , RX7.
% Tabela convertida para o formato do Matlab.
%
% Abaixo, atribuição valores de 'input' em Rx1, Rx2, ...,Rx7
for k = 1:max(size(Array))
    x = int2str(Array(k));
    name = ['Rx',x]; % Rever. Mesma coisa que o 'input'
    assignin('base',char(name),input.input{ 1,Array(k)})
    % Variáveis Rx1~Rx7 recebem valores de input
end
%
clear k x % Limpando a memória. Variáveis não são mais necessárias.
clear name % Limpando a memória. Variáveis não são mais necessárias.
%
% Definição das dimensões e vetores auxiliares de busca.
trainingProportion = 50; % Proporção definida como 50%. Variável 'trainingLimitNearIndex'
referenceTestProportion = 50; % Proporção da base usada para testes de referência
%

```

```

gridInputDimensions = size(input.input{1}); % = [36 2]. Matriz [36 linhas , 2 colunas]
samplesLength = gridInputDimensions(1); % = 36
%
trainingLimitNearIndex = samplesLength*trainingProportion/100; % = 18 (metade de 36)
trainingLimit = floor(trainingLimitNearIndex); % Round toward negative infinity = 18
% Arredondar em direção ao infinito negativo
%
% Definição da base usada para treinamento:
referenceTestLimitNearIndex = (samplesLength-trainingLimit)*referenceTestProportion/100; % = 4,5
% referenceTestLimitNearIndex = (36 - 18)*25/100 = 4,5
referenceTestLimit = floor(referenceTestLimitNearIndex); % 4,5. Round toward negative infinity = 4
% Arredondar em direção ao infinito negativo
%
logLikelihoodRegion = -160:0.2:0;
%
k1 = 1;
k4 = 1;
k5 = 1;
k6 = 1;
%
for k=1:LengthBakisTopology
    if k <= LengthBakisTopology - 1
        x = int2str(ArrayBakisTopology(k));
        x2 = int2str(ArrayBakisTopology(k+1));
        x3 = int2str(k);
        name = ['Rx',x,'ToRx',x2,'data_',x3]; % Cria str 'Rx?ToRx?data_?'
        name2 = [x,'-',x2];
        RegionMatrix_name(k,:) = name2;
        %ArrayBakis_RegionMatrix_1(k,:) = name; % Cria str Rx? e abaixo var Rx?
        %
        randomGridPostionIndex1 = randperm(samplesLength);
        % Random permutation. 1 ~ 36 (Permutação aleatória)
        % Vetor com 36 números aleatórios entre 1 e 36
        randomGridPostionIndex2 = randperm(samplesLength);
        % Random permutation. 1 ~ 36 (Permutação aleatória)
        % Vetor com 36 números aleatórios entre 1 e 36
        %
        for i = 1:samplesLength % samplesLength = 36
            %
            teste = input{str2double(x)}(randomGridPostionIndex2(i,:));
            teste2 = input{str2double(x2)}(randomGridPostionIndex1(i,:));
            MatrixNameBakisZeros(:,i) = [teste(:);teste2(:)];
        end
        assignin('base',char(name),MatrixNameBakisZeros)
        % Assinala valores de Rx?ToRx?data para variável Rx?ToRx?data.
        % Iterativo até completar todas as regiões Rx? para Rx?
        data(:, :, k) = MatrixNameBakisZeros;
        % Todos os valores de todas as regiões estão em 'data'
        % name1 = ['hmm_',name,]; % Cria str 'hmm_Rx?ToRx?data'
        % teste3 = RSA_hmmBuilder(MatrixNameBakisZeros(:,1:floor(trainingLimit)), 2, 2, 20);
        % assignin('base',char(name1),teste3)
        % outputs(k) = RSA6_fillResults(name2, teste3, data, trainingLimit, referenceTestLimit, samplesLength,
        % logLikelihoodRegion, k, LengthBakisTopology, name2);
        % outputs(k) = RSA_fillResults('1-2', teste3, data, trainingLimit, referenceTestLimit, samplesLength,
        % logLikelihoodRegion);
        %
        data_training(:, :, k) = MatrixNameBakisZeros(:, 1:trainingLimit);
        % data_training é parte usada para treinamento do HMM
        data_testing(:, :, k) = MatrixNameBakisZeros(:, trainingLimit+1:samplesLength);
        % data_testing é parte usada para teste
    end
end

```



```

%
for k2 = 1:trainingLimit
    data_training_all(:,k1) = MatrixNameBakisZeros(:,k2);
    k1 = k1+1;
end
%
% for k3 = trainingLimit+1:samplesLength
%   data_testing_all(:,k4) = MatrixNameBakisZeros(:,k3);
%   k4 = k4+1;
% end
%
for k3 = trainingLimit+1:samplesLength
    data_testing_all(:,k4) = MatrixNameBakisZeros(:,k3);
    %
    if k3 <= 27 % Data Test de 19 até 27 para limiar
        data_testing_all_limiar(:,k5) = MatrixNameBakisZeros(:,k3);
        k5 = k5+1;
    else % Data Test de 28 até 36 para teste de HMM e comparação limiar
        data_testing_all_test(:,k6) = MatrixNameBakisZeros(:,k3);
        k6 = k6+1;
    end
    k4 = k4+1;
end
end
%
end
%
nameAll = '/';
for k = 1:length(RegionMatrix_name)
    name3 = [RegionMatrix_name(k,:), '/'];
    nameAll = strcat(nameAll, name3); % Concatenate strings horizontally
end
%
hmm = RSA10_hmmBuilder(data_training_all(:,,:), 10, LengthBakisTopology, 200);
outputs = RSA10_fillResults...
(ArrayBakisTopology, hmm, data_testing_all, trainingLimit, referenceTestLimit,...
samplesLength, logLikelihoodRegion, k, LengthBakisTopology,...
RegionMatrix_name, data_testing_all_limiar, data_testing_all_test, nameAll);
% matrix
%
```

- `hmmBuilder.m`

```

% Matriz de Probabilidade de Estados é gerada por meio da rotina chamada "hmmBuilder.m"
function [hmm]= RSA10_hmmBuilder(data, M, Q, int)
%
% Entradas:
% data = arquivo de mffccs a ser utilizado para treinar a cadeia markoviana
% data = Rx1ToRx2data(:,1:floor(trainingLimit)) = Rx1ToRx2data(:,1 ~ 18)
% Limite de Treinamento = 18. Então pega todos os valores de 1 até 18.
% M = 2 = N° de misturas gaussianas nas probabilidades de missão de observáveis por estado
% Q = 2 = N° de Estados
% int = 20 = N° de iterações do treinamento
%
% Saída struct hmm com os seguintes campos:
% hmm.prior(i) = Pr(Q(1) = i) % Vetor de probabilidade inicial de estados
%
% hmm.transmat(i,j) = Pr(Q(t+1)=j | Q(t)=i) % Matriz de probabilidade de transição
```

```

% hmm.mu(:,j,k) = E[Y(t) | Q(t)=j, M(t)=k ] % Parâmetros média e desvio padrão
% hmm.Sigma(:,j,k) = Cov[Y(t) | Q(t)=j, M(t)=k] % das misturas gaussianas % (funções
% densidade de probabilidade de emissão
% de observáveis (MFCC)
% hmm.mixmat(j,k) = Pr(M(t)=k | Q(t)=j) Matriz de pesos das misturas gaussianas
%
% Etapa de Inicialização (variáveis fixas definidas acima)
%
si = size(data); % si = [2, 2, 36, 14]
% N° de iterações 2x2 em 36 sequências de observáveis em 14 regiões
% Parâmetros
O = si(1); % Número de janelas por amostra = 2
T = si(2); % Número de parâmetros observáveis por janela (MFCC) = 2
nex = si(3); % Número de amostras (sequências de observáveis) = 36
cov_type = 'full';
flag = 1; % Indicador de atendimento de critério de parada;
while flag == 1;
%
% Inicialização dos parâmetros
% Análise de agrupamento (clustering) - k-means com 'mixgauss_init'
%
[hmm.mu, hmm.Sigma] = RSA_mixgauss_init(... %
    Q*M, reshape(data, [O T*nex]), cov_type); % Inicialização das misturas
hmm.mu = reshape(hmm.mu, [O Q M]); % gaussianas (k-means)
hmm.Sigma = reshape(hmm.Sigma, [O O Q M]); %
hmm.mixmat = (1/M)*ones(Q,M); % Inicialização da matriz de pesos das
% misturas gaussianas
hmm.prior = [1 zeros(1,(Q-1))]; % Inicialização do vetor de probabilidade
% inicial de estados

hmm.transmat = zeros(Q,Q);
for l = 1:Q
    hmm.transmat(l,l:Q) = 1/(Q-l+1);
end
%
% Etapa de treinamento do HMM (cálculo com iterações)
[hmm.LLcomando, hmm.prior, hmm.transmat, hmm.mu, hmm.Sigma, hmm.mixmat, flag] = ...
    RSA_mhmm_emedit(data, hmm.prior, hmm.transmat, hmm.mu, hmm.Sigma, hmm.mixmat, 'max_iter',
int);
%
end

```

- fillResults.m

```

%
function [output] = RSA10_fillResults...
    (ArrayBakisTopology,hmm,data_testing_all,trainingLimit,...
    referenceTestLimit, samplesLength, logLikelihoodRegion, k,...
    LengthBakisTopology, RegionMatrix_name,...
    data_testing_all_limiar, data_testing_all_test,...
    nameAll)
%
output.selectedRegionName = ArrayBakisTopology;
regionIndexNames = nameAll;
%
% ----- Cálculo da Log-verossimilhanças -----
% for regionIndex = 1:LengthBakisTopology-1
%     selectedRegionIndex = find(selectedRegionName == regionIndexNames(regionIndex));
logLikes_limiar = zeros(1,length(data_testing_all_limiar));
logLikes_test = zeros(1,length(data_testing_all_test));

```

```

%
for dt = 1:length(data_testing_all_limiar)
    logLikes_limiar(:,dt) = RSA_calcLikelihood(data_testing_all_limiar(:,dt), hmm);
    % 'calcLikelihood.m' calcula log-verossimilhança p/ amostras de teste
    % Pega amostras de teste-referencia e calcula log-verossimilhança
end
%
for dt = 1:length(data_testing_all_test)
    logLikes_test(:,dt) = RSA_calcLikelihood(data_testing_all_test(:,dt), hmm);
    % 'calcLikelihood.m' calcula log-verossimilhança p/ amostras de teste
    % Pega amostras de teste-teste e calcula log-verossimilhança
end
%
% ----- Retirada de Outliers -----
% Amostras de teste de referência = selectedLogLikes_limiar = output.referenceLogLikes
% Amostras de teste de validação = selectedLogLikes_test = output.testLogLikes
%
allSortedLogLikes_limiar = sort(logLikes_limiar);
allSortedLogLikes_test = sort(logLikes_test);
%
randPermTest_limiar = randperm(length(data_testing_all_limiar)-2);
randPermTest_test = randperm(length(data_testing_all_test)-2);
%
SortedLogLikes_limiar = allSortedLogLikes_limiar(3:length(allSortedLogLikes_limiar));
SortedLogLikes_test = allSortedLogLikes_test(3:length(allSortedLogLikes_limiar));
%
selectedLogLikes_limiar = SortedLogLikes_limiar(randPermTest_limiar);
selectedLogLikes_test = SortedLogLikes_test(randPermTest_test);
%
output.referenceLogLikes = selectedLogLikes_limiar;
% Cria vetor de dados de Referência
output.testLogLikes = selectedLogLikes_test;
% Cria vetor de dados de Teste
% ksdensity = Estima fdp para vetores de Verossimilhanças.
% fdp de cada região será plotada em gráfico.
output.referenceDensities = ksdensity(output.referenceLogLikes, logLikelihoodRegion);
% Estimativa pdf do vetor de Referência (probability density estimate)
output.testDensities = ksdensity(output.testLogLikes, logLikelihoodRegion);
% Estimativa pdf do vetor de Teste (probability density estimate)
%
% ----- Cálculo de Limiars -----
%
% ----- Limiar 1 (Média das médias) -----
output.limiar1 = mean(output.referenceLogLikes);
limiar1PlotVector = zeros(1,length(logLikelihoodRegion));
limiar1Index = RSA_findIndex(output.limiar1, logLikelihoodRegion);
limiar1PlotVector(limiar1Index) = 1;
%
% ----- Limiar 2 (Média Cruzada) -----
output.limiar2 = (mean(output.referenceLogLikes)*std(output.referenceLogLikes));
limiar2PlotVector = zeros(1,length(logLikelihoodRegion));
limiar2Index = RSA_findIndex(output.limiar2, logLikelihoodRegion);
limiar2PlotVector(limiar2Index) = 1;
%
% ----- Limiar 3 (Percentil P20 das Log-Verossimilhanças) -----
output.limiar3 = prctile(output.referenceLogLikes,20); % Uso de Ref Limiar
limiar3PlotVector = zeros(1,length(logLikelihoodRegion));
limiar3Index = RSA_findIndex(output.limiar3, logLikelihoodRegion);
limiar3PlotVector(limiar3Index) = 1;
%

```

```

% ----- Limiar 4 (Percentil P20 das f.d.p) -----
nonZeroDensityIndices = find(output.referenceDensities);
% find = Find indices and values of nonzero elements
% prctile = Percentiles of a data set
output.limiar4 = prctile(output.referenceDensities(nonZeroDensityIndices),20);
% output.limiar4 = prctile(output.referenceDensities,20);
limiar4PlotVector = zeros(1,length(logLikelihoodRegion));
limiar4Index = RSA_findIndex(output.limiar4,logLikelihoodRegion);
limiar4PlotVector(limiar4Index) = 1;
%
% output.limiar1FalsePositives = 0;
output.limiar1FalseNegatives = 0;
output.limiar1TruePositives = 0;
% output.limiar1TrueNegatives = 0;
% %
% output.limiar2FalsePositives = 0;
output.limiar2FalseNegatives = 0;
output.limiar2TruePositives = 0;
% output.limiar2TrueNegatives = 0;
%
% output.limiar3FalsePositives = 0;
output.limiar3FalseNegatives = 0;
output.limiar3TruePositives = 0;
% output.limiar3TrueNegatives = 0;
%
% output.limiar4FalsePositives = 0;
output.limiar4FalseNegatives = 0;
output.limiar4TruePositives = 0;
% output.limiar4TrueNegatives = 0;
%
for k = 1:length(output.testLogLikes)
    if output.testLogLikes(k) < output.limiar1
        output.limiar1FalseNegatives = output.limiar1FalseNegatives+1;
    else
        output.limiar1TruePositives = output.limiar1TruePositives+1;
    end
    %
    if output.testLogLikes(k) < output.limiar2
        output.limiar2FalseNegatives = output.limiar2FalseNegatives+1;
    else
        output.limiar2TruePositives = output.limiar2TruePositives+1;
    end
    %
    if output.testLogLikes(k) < output.limiar3
        output.limiar3FalseNegatives = output.limiar3FalseNegatives+1;
    else
        output.limiar3TruePositives = output.limiar3TruePositives+1;
    end
    %
    if output.testLogLikes(k) < output.limiar4
        output.limiar4FalseNegatives = output.limiar4FalseNegatives+1;
    else
        output.limiar4TruePositives = output.limiar4TruePositives+1;
    end
end
%
disp(strcat('Score do HMM das Regiões: ',regionIndexNames));
%
disp(' Limiar 1 (Média das médias): ');

```

```

disp(strcat(' Verdadeiros Positivos:
',num2str(output.limiar1TruePositives),'/',num2str(length(output.testLogLikes))));
disp(strcat(' Falsos Negativos:
',num2str(output.limiar1FalseNegatives),'/',num2str(length(output.testLogLikes))));
%
disp(' Limiar 2 (Média Cruzada): ');
disp(strcat(' Verdadeiros Positivos:
',num2str(output.limiar2TruePositives),'/',num2str(length(output.testLogLikes))));
disp(strcat(' Falsos Negativos:
',num2str(output.limiar2FalseNegatives),'/',num2str(length(output.testLogLikes))));
%
disp(' Limiar 3 (Percentil P20 das Log-Verossimilhanças): ');
disp(strcat(' Verdadeiros Positivos:
',num2str(output.limiar3TruePositives),'/',num2str(length(output.testLogLikes))));
disp(strcat(' Falsos Negativos:
',num2str(output.limiar3FalseNegatives),'/',num2str(length(output.testLogLikes))));
%
disp(' Limiar 4 (Percentil P20 das f.d.p): ');
disp(strcat(' Verdadeiros Positivos:
',num2str(output.limiar4TruePositives),'/',num2str(length(output.testLogLikes))));
disp(strcat(' Falsos Negativos:
',num2str(output.limiar4FalseNegatives),'/',num2str(length(output.testLogLikes))));
disp(' ');
%
figure
%
hold on
%
hall = plot...
    (...
    logLikelihoodRegion, output.testDensities, 'r',...
    logLikelihoodRegion,limiar1PlotVector, 'y',...
    logLikelihoodRegion,limiar2PlotVector, 'm',...
    logLikelihoodRegion,limiar3PlotVector, 'g',...
    logLikelihoodRegion,limiar4PlotVector, 'c'...
    );
%
xlim([-60 0])
ylim([0 0.5])
legend({'f.d.p. vetor log-verossimilhança de teste',...
    'Limiar1 (Média das Log-Verossimilhanças)',...
    'Limiar2 (Média Cruzada)',...
    'Limiar3 (Percentil P20 Log-Verossimilhanças)',...
    'Limiar4 (Percentil P20 f.d.p)'},...
    'FontSize',18,'Location','northwest');
legend('boxoff')
%
title({'Regiões percorridas: ',regionIndexNames},'FontSize',18,'FontWeight','bold');
hall(1).LineWidth = 2.0;
xlabel({'Log-Verossimilhanças'},'FontSize',20);
ylabel({'Densidade de Probabilidades'},'FontSize',20);
hold off
end
%
```