

Universidade de São Paulo
Instituto de Física

SUVREL aplicado em comparação de voz para fins forenses no arcabouço bayesiano e aplicabilidade do Teorema de Bayes no ordenamento jurídico brasileiro

(versão corrigida)

Eduardo Rodrigues da Silva

Orientador: Prof. Dr. Manfredo Harri Tabacniks

Tese de doutorado apresentada ao Instituto de Física da
Universidade de São Paulo como requisito parcial para a
obtenção do título de Doutor em Ciências.

Banca Examinadora:

Prof. Dr. Manfredo Harri Tabacniks (IFUSP)
Prof. Dr. Nestor Felipe Caticha Alfonso (IFUSP)
Prof. Dr. Miguel Arjona Ramírez (EPUSP)
Dr. Fabio Henrique Jagosich (SPTC / Instituto de Criminalística)
Prof. Dr. Geoffrey Stewart Morrison (Aston University / UK)

São Paulo
2021

Autorizo a reprodução e divulgação total ou parcial deste trabalho, por qualquer meio convencional ou eletrônico, para fins de estudo e pesquisa, desde que citada a fonte.

FICHA CATALOGRÁFICA
Preparada pelo Serviço de Biblioteca e Informação
do Instituto de Física da Universidade de São Paulo

Silva, Eduardo Rodrigues da

SUVREL aplicado em comparação de voz para fins forenses no arcabouço bayesiano e aplicabilidade do Teorema de Bayes no ordenamento jurídico brasileiro. São Paulo, 2021.

Tese (Doutorado) — Universidade de São Paulo. Instituto de Física.

Orientador: Prof. Dr. Manfredo Harri Tabacniks.

Departamento: Física Aplicada.

Área de Concentração: Sistemas Complexos.

Unitermos: 1. Física Matemática; 2. Processamento de Sinais Acústicos; 3. Física Forense; 4. Inferência Bayesiana; 5. SUVREL.

USP/IF/SBI-033/2021

University of São Paulo
Physics Institute

SUVREL applied in voice comparison for
forensic purposes in the Bayesian framework
and applicability of the Bayes' Theorem in the
Brazilian legal system

(corrected version)

Eduardo Rodrigues da Silva

Supervisor: Prof. Dr. Manfredo Harri Tabacniks

Thesis submitted to the Physics Institute of the University
of São Paulo in partial fulfillment of the requirements for
the degree of Doctor of Science.

Examining Committee:

Prof. Dr. Manfredo Harri Tabacniks (IFUSP)
Prof. Dr. Nestor Felipe Caticha Alfonso (IFUSP)
Prof. Dr. Miguel Arjona Ramírez (EPUSP)
Dr. Fabio Henrique Jagosich (SPTC / Instituto de Criminalística)
Prof. Dr. Geoffrey Stewart Morrison (Aston University / UK)

São Paulo
2021

*Para Cassia, meu amor, que me apoiou em cada passo dessa jornada.
Para Genilda, minha mãe, que me ensinou a ler e a gostar de estudar.
Para minha filha Ingrid, minha amiga e meu orgulho.
Para minha maninha Andréia, “sindomididown” mais linda do mundo, que, do alto
de sua simplicidade para perceber a vida, sempre tem uma palavra de apoio ao
maninho “atapalhado”.*

Agradecimentos

Por todo apoio e incentivo, gostaria de agradecer:

Ao meu orientador, **Prof. Dr. Manfredo Harri Tabacniks**, pelo apoio e confiança inabaláveis durante todos esses anos e por ser o meu modelo do que espero ser quando for a minha vez de orientar.

Ao meu amigo – de fato, irmão de pais diferentes – **André Borges** por todas as longas discussões e por me auxiliar nos aspectos jurídicos que essa tese tomou.

À minha amiga **Edna** por ter me feito enxergar luz no momento mais escuro. Saiba que serei eternamente grato a você.

Ao **Rinaldo** e à **Selma** por terem me impulsionado nos momentos em que mais precisei.

À **Promotora de Justiça Dra. Eliana Faleiros Vendramini Carneiro** com quem tive a primeira conversa de onde surgiu a centelha que inspirou essa tese.

Ao **Prof. Dr. Zwinglio de Oliveira Guimarães Filho**, por ter me auxiliado nas inúmeras dúvidas que tive em Matlab.

Ao **Prof. Dr. Josif Frenkel** pela demonstração de humanidade e compreensão que sempre me inspirarão a tentar um dia chegar à metade da pessoa que ele é.

Ao **Dr. Jonatas Cesar** e ao **Prof. Dr. Nestor Felipe Caticha Alfonso** por gentilmente esclarecerem minhas dúvidas sobre o SUVREL.

À **Promotora de Justiça Dra. Sandra Rodrigues De Oliveira M. Barbuto** por ter sido minha primeira “cobaia” quando prontamente cedeu seus textos para meus primeiros testes de análise comunicativa.

Às minhas primeiras “cobaias” que cederam suas vozes para meus testes iniciais, saibam que vocês foram muito importantes nesse caminhar: **Edna, Cassia, Diego, Humberto, Paula, Regina, Renata, Fabrício, Fumach, Leandro, Mauro, Rinaldo, Selma, Agostinho, Alexandre, Bueno, Fernandinho, Luiz, Marcelo Fernandes, Marco Antonio, Marianno, Reinaldo, Renato, Richard, Santana, Silvio e Vicente.** Peço desculpas se deixei de citar alguém.

Agradeço também à **Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES)** pelo suporte financeiro quando da apresentação do artigo “Supervised Variational Relevance Learning (SUVREL) applied to voice comparison” na “17^a Conferência Internacional Australasiana sobre Ciência e Tecnologia da Fala” em Sydney, Austrália.

RESUMO

Trata-se de tese que estudou a Identificação Positiva de Falantes para fins forenses. Foram escolhidas duas abordagens distintas: análise comunicativa e análise de voz. A análise comunicativa consiste na aplicação de recursos da Estilometria e da Linguística para classificação por pertencimento de um indivíduo a determinado grupo através do estudo de seu vocabulário, essencialmente, em material escrito – textos, e-mails – ou transcritos – interceptações telefônicas judiciais. A análise de voz consiste em desenvolver ferramental estatístico para identificação de um indivíduo com base em parâmetros acústicos extraídos de gravações de voz que representem, por exemplo, frequências ressonantes provenientes de traços da conformação físico-anatômica do falante, bem como, de seus hábitos de fala – idioma, dialetos, sotaques etc. Essa etapa associou, de forma inédita, as técnicas usuais de obtenção de parâmetros de voz com um método de aprendizado por relevância, o *Supervised Variational Relevance Learning* (SUVREL). Em análise comunicativa, buscou-se associar a razão de distinção com a base de referência do número de resultados de pesquisas no *Google*, mas uma barreira metodológica resultou em estimativas enviesadas. Na análise de voz, foram feitos três experimentos de separação automática de dados via Análise de Componentes Principais, antes e depois do pré-processamento com SUVREL. Os resultados iniciais mostraram significativo aumento na concentração da variância total explicada pelos três primeiros componentes principais, logo, indicativo de benefício no emprego do SUVREL em análise de voz. Definindo-se a similaridade entre vozes como sendo a superposição volumétrica dos escores no espaço PCA, um experimento com vozes de 100 homens resultou em decréscimo percentual médio de 20.44 ± 0.05 %. No contexto pericial forense, realizou-se abordagem inovadora, no âmbito do sistema judicial brasileiro, ao fazer-se a extensão da discussão sobre conciliar o preceito da *presunção da não culpabilidade* com a aplicação do Teorema de Bayes. Concluiu-se que, de modo semelhante às discussões em outros países, onde vigora a *presunção de inocência*, o cerne recai sobre o mesmo verbo epistêmico *presumir* e apenas a presença da dúvida mínima *a priori* contida nesse componente semântico já é condição suficiente para a validade da aplicação jurídica do Teorema de Bayes no contexto do ordenamento jurídico brasileiro.

PALAVRAS-CHAVE: Física Matemática; Processamento de Sinais Acústicos; Física Forense; Inferência Bayesiana; SUVREL.

ABSTRACT

This thesis investigates the Positive Identification of Speakers for forensic purposes. Two different approaches were chosen: communicative analysis and voice analysis. The communicative analysis consists of the application of Stylometry and Linguistics resources to classify individuals as belonging to a certain group through the study of their vocabulary, essentially in written material – texts, e-mails – or transcripts – judicial telephone interceptions. Voice analysis consists of developing statistical tools to identify an individual based on acoustic parameters extracted from voice recordings that represent, for example, resonant frequencies from traits of the speaker's physical-anatomical conformation, as well as his speaking habits – language, dialects, accents etc. This step associated, in an unprecedented way, the usual techniques for obtaining voice parameters with a learning method by relevance, the *Supervised Variational Relevance Learning* (SUVREL). In communicative analysis, we sought to associate the distinctiveness ratio with the reference base of the number of search results on *Google*, but a methodological barrier resulted in biased estimates. In the voice analysis, three experiments of automatic data separation via Principal Component Analysis were carried out, before and after the pre-processing with SUVREL. The initial results showed a significant increase in the concentration of total variance explained by the first three main components, therefore, indicative of benefit in the use of SUVREL in voice analysis. Defining the similarity between voices as the volumetric superposition of scores in the PCA space, an experiment with voices of 100 men resulted in an average percentage decrease of $20.44 \pm 0.05 \%$. In the forensic expert context, an innovative approach was taken, within the scope of the Brazilian judicial system, by extending the discussion on reconciling the precept of *presumption of non-culpability* with the application of Bayes' Theorem. It was concluded that, similarly to discussions in other countries, where the *presumption of innocence* prevails, the core lies on the same epistemic verb *to presume* and only the presence of the minimal doubt *a priori* contained in this semantic component is already a sufficient condition for the validity of the legal application of Bayes' Theorem in the context of the Brazilian legal system.

KEYWORDS: Mathematical Physics; Acoustic Signal Processing; Forensic Physics; Bayesian Inference; SUVREL.

Sumário

	página
1. Introdução	15
2. Aspectos gerais da perícia criminal no ordenamento jurídico brasileiro e aplicação do Teorema de Bayes para a apresentação de seus resultados	17
2.1. Breve discussão sobre os aspectos jurídicos	20
2.1.1. Prova e livre convencimento	20
2.1.2. Valoração da prova pericial	22
2.1.3. Presunção da não culpabilidade ou presunção de inocência? . .	24
2.1.3.1. Aspectos linguísticos da <i>presunção</i> de inocência e sua relação com o Teorema de Bayes	26
2.1.4. Aceitabilidade Internacional do Teorema de Bayes na apresenta- ção de resultados periciais	29
2.2. Vieses cognitivos: uma visão geral	32
2.2.1. “Paradoxo de Monty Hall” e a ilusão do controle	39
2.2.1.1. Dissonância cognitiva	41
2.2.1.2. Efeito Intercultural	41
2.2.1.3. Comparação entre espécies: humanos x pombos	42
2.3. Teorema de Bayes	44
2.3.1. Conceitos fundamentais	46
2.3.1.1. Independência	46
2.3.1.2. Probabilidade Conjunta	47

2.3.1.3.	Permutabilidade	47
2.3.1.4.	Regra da Probabilidade Total	48
2.3.2.	Formulação do Teorema de Bayes	49
2.3.3.	Desvendando o Dilema de Monty Hall usando o Teorema de Bayes	52
2.3.4.	Aplicação do Teorema de Bayes no contexto epidemiológico . .	54
2.3.4.1.	CASO 1: interpretação de exame em indivíduo de sub- grupo populacional COM sintomas ou COM histórico de exposição ao vírus Sars-CoV-2	55
2.3.4.2.	CASO 2: interpretação de exame em indivíduo de sub- grupo populacional SEM sintomas ou SEM histórico de exposição ao vírus Sars-CoV-2	58
2.3.4.2.1.	Obtenção do índice de falsos negativos	59
2.3.4.2.2.	Obtenção do índice de falsos positivos	60
2.3.4.2.3.	Aplicação recursiva do Teorema de Bayes	62
2.3.5.	Razão de Verossimilhanças	64
2.3.5.1.	Aplicação recursiva da Razão de Verossimilhanças . .	66
2.3.5.2.	Abordagem gráfica para obtenção da Razão de Verossi- milhanças	67
2.3.5.3.	Apresentação dos resultados: LR ou $\log LR$	70
2.3.5.4.	Incertezas	71
2.4.	Abordagem bayesiana para apresentação de resultados periciais	72
2.4.1.	Exemplo no contexto de análise de voz	73
2.4.2.	Exemplo: comparação entre objetos apreendidos e filmagens de um crime	74

3. Identificação de falantes	85
3.1. Som, voz, fala e abordagens forenses	86
3.1.1. Som e audição humana	86
3.1.2. A produção da voz	89
3.1.2.1. Articuladores Vocais	91
3.1.3. Reconhecimento de falantes	92
3.1.3.1. Fontes de Variabilidade no Reconhecimento de Falantes	94
3.1.3.1.1. Parâmetros Acústicos	96
3.1.3.2. Caracterização de Falantes	98
3.1.3.3. Abordagens na Identificação Forense de Falantes	99
3.1.3.4. Biometria	102
3.1.3.5. Disfarce da Voz	103
3.2. Processamento de sinais	103
3.2.1. Diferenças básicas entre sinais de tempo contínuo e de tempo discreto	105
3.2.2. Série de Fourier	108
3.2.3. Transformada de Fourier em Sinais de Tempo Contínuo	110
3.2.4. Transformada de Fourier em Sinais de Tempo Discreto	111
3.2.5. Teorema da Amostragem	112
3.2.6. Representações de sinais contínuos e discretos	113
3.2.7. Sistemas Digitais	114
3.2.8. Operador Convolução e Sistemas Lineares Invariantes no Tempo	115
3.3. Modelos para a voz	120

3.3.1.	A escala mel	121
3.3.1.1.	Coeficientes Cepstrais de Frequência Mel (MFCC)	122
3.3.1.2.	Deltas e delta-deltas	123
3.3.1.3.	Banco de filtros Mel	124
3.4.	Análise de Componentes Principais	127
3.4.1.	Obtenção dos componentes principais	128
3.5.	SUVREL: Supervised Variational Relevance Learning	132
4.	Objetivos Gerais	137
5.	Desenvolvimento e experimentos	139
5.1.	Análise Comunicativa	141
5.1.1.	Razão de distinção	141
5.1.2.	Cálculo da razão de distinção empregando resultados do <i>Google</i>	142
5.2.	Análise de Voz	144
5.2.1.	Experimento preliminar	144
5.2.1.1.	Métodos	144
5.2.1.2.	Resultados	145
5.2.1.3.	Conclusões	145
5.2.2.	Experimento intermediário: Supervised Variational Relevance Learning (SUVREL) aplicado à comparação de voz	149
5.2.2.1.	Introdução	149
5.2.2.2.	Métodos	150
5.2.2.3.	Resultados	151
5.2.2.4.	Conclusões	159

5.2.3. Experimento final	159
5.2.3.1. Métodos	159
5.2.3.2. Resultados	161
5.2.3.3. Conclusões	167
6. Conclusões	169
6.1. Abordagem por análise comunicativa	170
6.2. Abordagem por análise de voz	170
6.2.1. Aprofundamentos futuros	172
6.3. Teorema de Bayes na apresentação de Resultados Periciais e sua relação com o Ordenamento Jurídico Brasileiro	172
6.4. Contribuições científicas inovadoras	174
6.5. Impacto social	175
Referências Bibliográficas	177
Apêndices	185
A. Gráficos	187
A.1. UBM: 228 homens / frequência: 8kHz / qualidade original	187
A.2. UBM: 228 homens / frequência: 8kHz / SNR: 25dB	198
A.3. UBM: 319 mulheres / frequência: 8kHz / qualidade original	209
A.4. UBM: 319 mulheres / frequência: 8kHz / SNR: 25dB	220
B. Códigos Matlab	231
B.1. Script “MFCC_PCA_script_intravariability.m”	231
B.2. Script “MFCC_PCA_script_UBM_SUVREL.m”	251

B.3. Função “SUVREL.m” 281

B.4. Função “ellipsoidVolumeComparison.m” 283

B.5. Função “ellipsoidVolumeComparisonErrorTest.m” 286

B.6. Função “melTriangularFilterBank.m” 288

B.7. Função “freq2melscale.m” 290

B.8. Função “melscale2freq.m” 290

1. Introdução

Como servidor do Ministério Público do Estado de São Paulo, mais especificamente Oficial de Promotoria Chefe do Grupo de Atuação Especial de Combate ao Crime Organizado, Núcleo São Paulo, Subnúcleo Capital, entre os anos de 2008 e 2015, tive a oportunidade de travar contato com vários processos em que os Promotores de Justiça faziam uso de interceptações telefônicas, autorizadas judicialmente em medidas cautelares, como elemento de formação de prova e linha norteadora de investigações. Esse contato, ainda que superficial, despertou meu interesse pelo campo das ciências forenses.

Atualmente, atuando como Perito Criminal do Instituto de Criminalística do Estado de São Paulo, especificamente no Núcleo de Identificação Criminal, posso acompanhar em primeira mão esses casos.

Nas investigações em que se faz uso desse recurso, não é incomum que haja a necessidade de se validar, por métodos de comparação, as identidades dos falantes. Tanto do ponto de vista da defesa técnica do réu, que pode alegar não ser o autor daquele conteúdo, quanto da posição da promotoria, que tem por dever institucional promover a justiça e precisa certificar-se de que o material de áudio analisado realmente relaciona-se ao réu.

A busca pela automatização dos métodos de comparação e identificação positiva de falantes é tema atual. Como exemplo, as empresas *Nuance Communications* e *Voice Vault* possuem cerca de 46 milhões de pessoas catalogadas em sistemas de

biometria de voz ou soluções que usem impressões vocais para assegurar a prevenção de fraudes (MASTERSON, 2012). A biometria por voz não só tem assegurada seu papel na perícia criminal como também na indústria de telecomunicações, na área da saúde e finanças.

Dessa forma, o presente texto tem como motivação ampliar os métodos de análise de voz, fala e linguagem conhecidos, buscar aumentar os níveis de confiabilidade das técnicas atuais de identificação biométrica, desenvolver algoritmos que permitam automatizar essa tarefa e contribuir socialmente, posto que é de interesse público a criação de metodologias cada vez mais eficazes para a identificação positiva de falantes. Especificamente, neste trabalho, foi investigado o impacto do SUVREL (Supervised Variational Relevance Learning) na separação automática de vozes.

Trata-se de proposta de caráter multidisciplinar envolvendo, a princípio, Fonética, Linguística, Física Forense e Computação.

Considerando minha formação científica em Exatas, tem ficado evidente a deficiência de cursos/matérias voltadas aos métodos estatísticos em análise pericial e para apresentação de resultados e conclusões periciais nas demais formações de nível superior. O cargo de Perito Criminal exige formações superiores nos mais diversos campos do conhecimento e, dada a característica autocontida adotada pela grande maioria dos cursos superiores do Brasil – onde a interdisciplinariedade é exceção – não é de se esperar que estatística para apresentação de resultados seja conhecimento comum à categoria dos Peritos Criminais.

Os resultados obtidos no presente trabalho foram apresentados em termos de Razão de Verossimilhança (Teorema de Bayes). Foi incluída uma ampla seção em que são explicados a origem, a necessidade e exemplos de aplicação do Teorema de Bayes. Na referida seção, também foram abordados os aspectos jurídicos que fundamentam e validam o uso do Teorema de Bayes no contexto judicial e pericial.

2. Aspectos gerais da perícia criminal no ordenamento jurídico brasileiro e aplicação do Teorema de Bayes para a apresentação de seus resultados

Sumário

2.1. Breve discussão sobre os aspectos jurídicos	20
2.1.1. Prova e livre convencimento	20
2.1.2. Valoração da prova pericial	22
2.1.3. Presunção da não culpabilidade ou presunção de inocência?	24
2.1.3.1. Aspectos linguísticos da <i>presunção</i> de inocência e sua relação com o Teorema de Bayes	26
2.1.4. Aceitabilidade Internacional do Teorema de Bayes na apresentação de resultados periciais	29
2.2. Vieses cognitivos: uma visão geral	32
2.2.1. “Paradoxo de Monty Hall” e a ilusão do controle	39
2.2.1.1. Dissonância cognitiva	41
2.2.1.2. Efeito Intercultural	41
2.2.1.3. Comparação entre espécies: humanos x pombos	42

2.3. Teorema de Bayes	44
2.3.1. Conceitos fundamentais	46
2.3.1.1. Independência	46
2.3.1.2. Probabilidade Conjunta	47
2.3.1.3. Permutabilidade	47
2.3.1.4. Regra da Probabilidade Total	48
2.3.2. Formulação do Teorema de Bayes	49
2.3.3. Desvendando o Dilema de Monty Hall usando o Teorema de Bayes .	52
2.3.4. Aplicação do Teorema de Bayes no contexto epidemiológico	54
2.3.4.1. CASO 1: interpretação de exame em indivíduo de subgrupo populacional COM sintomas ou COM histórico de exposição ao vírus Sars-CoV-2	55
2.3.4.2. CASO 2: interpretação de exame em indivíduo de subgrupo populacional SEM sintomas ou SEM histórico de exposição ao vírus Sars-CoV-2	58
2.3.5. Razão de Verossimilhanças	64
2.3.5.1. Aplicação recursiva da Razão de Verossimilhanças	66
2.3.5.2. Abordagem gráfica para obtenção da Razão de Verossimilhanças	67
2.3.5.3. Apresentação dos resultados: LR ou $\log LR$	70
2.3.5.4. Incertezas	71
2.4. Abordagem bayesiana para apresentação de resultados periciais	72
2.4.1. Exemplo no contexto de análise de voz	73
2.4.2. Exemplo: comparação entre objetos apreendidos e filmagens de um crime	74

No campo das Ciências Forenses, a interdisciplinariedade é fortemente presente. Áreas do conhecimento humano, tais como Biologia, Física, Química, Medicina etc. exercidas por Peritos Criminais, Médicos Legistas e Odontologistas, aliam-se ao Direito,

à Criminologia e à Criminalística para auxiliar e orientar as partes componentes do sistema Judiciário para que a Justiça seja aplicada de modo parcimonioso e escoreito.

A atividade jurídica é exercida por humanos que são inerentemente influenciáveis e conduzidos, frequentemente, por ímpetos decisórios em que pesam – muitas vezes inconscientemente, portanto, ausentes elementos de intencionalidade – vieses cognitivos, dentre eles, a “Ilusão do Controle”

Vieses Cognitivos! ilusão de controle. Além disso, dada a estrutura autocontida da educação formal, vigente praticamente em todos os países, em que a interdisciplinariedade é posta de lado em prol de formações altamente especializadas voltadas ao mercado de trabalho, muitos profissionais desconhecem discussões ou soluções já consolidadas para os problemas diários que enfrentam – problemas estes que podem ainda sequer terem sido reconhecidos por eles mesmos como tais.

No presente texto, serão apresentados e discutidos, à luz do regramento jurídico brasileiro, meios já consolidados nas ciências exatas para conciliar o papel independente e isento dos Peritos Criminais, na apresentação dos resultados obtidos em seus exames, com uma avaliação jurídica isenta de – ou minimamente permeada por – vieses cognitivos. Também é detalhado o “ferramental” matemático baseado no Teorema de Bayes e na Razão de Verossimilhanças, que permitirá ao julgador sopesar as evidências juntadas aos processos criminais sem desconsiderar fatores facilmente negligenciáveis – tais como a consideração simultânea de taxas de falsos positivos e falsos negativos – em face de suas concepções pessoais sobre a força das demais evidências subjetivas levantadas no decurso processual.

Com a proposta de que este trabalho abarque tantos campos distintos, com conhecimentos basilares tão diversos, a abordagem será altamente explicativa e exemplificativa, mas também não deixando de lado os aspectos técnicos necessários.

2.1. Breve discussão sobre os aspectos jurídicos

O magistrado brasileiro, quando “dizendo o Direito” – sua atividade primária – deve ter alguns princípios em mente: entre eles a presunção da não culpabilidade e, principalmente, o devido processo legal. Deve também ter o cuidado de sopesar todas as informações que lhe são fornecidas (provas periciais, depoimentos e inclusive suas convicções pessoais) quando tomando suas decisões. Apesar da aparente liberdade que lhe daria o princípio do livre convencimento, vale lembrar que este deve ser **devidamente motivado**, como propõe Dworkin¹ (2007, p. 275 apud IRIBURE, JR.; ROSA, 2018):

O ato de decidir não pode ser visto como a possibilidade de os julgadores escreverem uma série de contos independentes, sem qualquer correlação entre si, mas, ao contrário, antes, devem ter o compromisso com um romance único, cujo enredo se mostre integrado.

Essas decisões devem, portanto, ser caracterizadas pela integridade sistêmica – sendo o julgador capaz de lidar de maneira efetiva e consciente com os diversos vieses cognitivos a que está sujeito.

2.1.1. Prova e livre convencimento

No ordenamento jurídico brasileiro, em especial nos chamados códigos processuais – Penal (BRASIL, 1941) e Civil (BRASIL, 2015), entre outros – figura de grande importância a ideia do livre convencimento do magistrado, baseado, entre outros, na prova.

A prova pericial, um dos tipos de prova admitidos segundo as leis vigentes, é peça importante na engrenagem processual não apenas no Brasil, como também em países com culturas jurídicas tão diversas como os Estados Unidos da América

¹DWORKIN, R. **O império do direito**. 2. ed. São Paulo: Martins Fontes, 2007. Trad.: Jefferson Luiz Camargo.

ou a nossa vizinha Argentina. Recentemente, uma importante mudança no Código de Processo Penal (o CPP) – que estabeleceu normativamente a chamada “cadeia de custódia” – veio a reforçar a importância desse tipo de prova nos procedimentos processuais, notadamente os penais, mas também — por analogia — nos cíveis.

Por essa razão e outras elencadas a seguir, este capítulo propõe-se a discorrer de um modo um pouco mais abrangente sobre a importância do procedimento dos peritos oficiais na produção e tratamento da prova pericial, em especial para as Ciências Criminais que, segundo Sala (2018):

...propõem procedimentos técnico-científicos para desenvolver aptidões e habilidades fundamentais à prática jurídica nas mais variadas áreas de atuação da Criminalística. Com aplicação de métodos de estudo na elucidação e ações conclusivas pela Medicina Legal e pela Criminalística a finalidade está direcionada na elucidação de delitos e ações criminosas.

Ainda segundo o autor, a perícia criminal se baseia, no Brasil, fortemente “na coleta e análise de vestígios e/ou evidências por peritos nos locais de delito, sendo a carência e inobservância de procedimentos um ato que afeta a integridade destes locais” (SALA, 2018), o que justificou, entre outras coisas, um cuidado especial na definição e normatização da já mencionada cadeia de custódia que veio proposta na recente lei 13.964 de 2019, chamada de “Pacote Anticrime” (BRASIL, 2019).

Pesquisa de artigos escritos relacionados ao tema permite entender que tem existido, recentemente, no bojo dessa “revitalização” da importância da prova pericial, investimentos de importância expressiva – efetuados pelo governo federal e por suas contrapartes em nível estadual – com o objetivo de modernizar os institutos e órgãos responsáveis pela perícia criminal. Isso, conjugado também com outras iniciativas no âmbito de inovações normativas torna possível “se esperar a produção de provas periciais com elevada qualidade, atendendo critérios de objetividade e credibilidade.” (LIMA NETO; ALBANO, 2020).

Autores como Tavares e Andrade (2013) relacionam a prova pericial com as

garantias constitucionais do devido processo legal, pois se tratam de uma forma importante por meio da qual a autoridade judicial pode ter conhecimento de informações relacionados aos fatos e à autoria dos delitos e, assim, formar sua convicção com vistas a um julgamento justo.

2.1.2. Valoração da prova pericial

Quanto a isso, também nos diz Soares (2020, p. 232) – em seu artigo sobre as provas periciais e as garantias constitucionais – que:

É na fase pré-processual que ocorrem os atos investigatórios e são produzidas praticamente todas as provas periciais. É nesse momento que emerge um dos maiores problemas contemporâneos de nossa legislação criminal, em especial a de caráter processual, que alcança o ápice no momento que o juiz utiliza provas produzidas fora do espectro do contraditório e da ampla defesa para fundamentar sentenças condenatórias. Destacam-se entre essas provas as periciais, que são reguladas pelo CPP/1941 e estão situadas topograficamente associadas ao inquérito policial, isto é, compõem, da forma que se encontram, a investigação policial.

Ainda considerando a valoração da prova pericial, vemos que autores, tais como Sousa (2016, p. 19), defendem que:

A prova pericial tem que ser apreciada pelo julgador a três níveis: (1) quanto à sua validade (respeitante à sua regularidade formal); (2) quanto à base de facto pressuposta na perícia e (3) quanto à própria conclusão da perícia. Com relação ao primeiro nível, o magistrado analisará se a perícia foi realizada de acordo com a lei e com normas aceitáveis.

No que tange ao primeiro nível, há que aferir se a prova foi produzida de acordo com a lei, se não foi produzida contra proibições legais e examinar se o procedimento da perícia está de acordo com normas da técnica ou da prática corrente.

Falando um pouco mais sobre o valor da prova pericial como fonte importante de informações para as partes envolvidas no processo, Figueiredo e Sampaio (2020)

nos ensinam que a Lei 13.964 de 2019, usualmente chamada de “Pacote Anticrime” (BRASIL, 2019) tratou, entre outras coisas – como já mencionado – da normatização de alguns procedimentos a ela relacionados (entre eles a cadeia de custódia e a valorização de cada tipo de prova). Entre outras coisas, eles também nos ensinam que

Impulsionadas pelos avanços científicos e tecnológicos, as provas periciais hoje ocupam um lugar de destaque e relevância dentro do processo criminal, possuindo em seus resultados um grau elevado de confiança pelo senso comum e também pelas autoridades públicas, o que acarreta problemas diante da ausência de uma perspectiva crítico-racional, principalmente por parte dos julgadores. A falta de controle da validade e confiabilidade e, por conseguinte, o ingresso no processo de dados incorretos ou exagerados conduzem ao aumento do risco de erro nas decisões judiciais (FIGUEIREDO; SAMPAIO, 2020, p. 29).

Por fim, como se vê na mídia, a própria perícia criminal — parte da investigação de crimes — caiu no gosto do público, cativado por seriados estadunidenses tais como os da franquia “CSI – Crime Scene Investigation” (ZUIKER, 2000-2015). Essa franquia se tornou tão difundida que virou referência: a assassina condenada Suzane von Richthofen e os coautores do assassinato de seus pais (irmãos Cravinhos) tomaram como referência a série de televisão quando planejaram o crime. Queriam burlar a posterior investigação e assim tentar escapar de maneira impune (CAMPBELL, 2020).

O assim chamado “CSI effect” — estudado por especialistas em Ciências Criminais – é algo tão real que afeta a própria percepção de jurados com relação às provas periciais apresentadas por peritos oficiais em julgamentos: o jurado tem a “sensação” de que a prova apresentada em juízo tem qualidade inferior àquela que poderia ter. Assim, inconscientemente, os jurados acabam por subvalorizar provas cruciais, levados por uma percepção subjetiva (MAHER, 2009) (COLE; DIOSO-VILLA, 2009).

Com todos esses aspectos sob escrutínio, parece relevante desenvolver este trabalho tratando sobre a prova pericial em especial no aspecto de sua qualidade, apresentação e inteligibilidade pelo público cientificamente leigo, que, em última instância,

compõe-se não apenas pelos membros do júri, como também pelas autoridades e agentes policiais, membros do ministério público, magistrados e outros operadores do direito. Com isso, reforça-se sua importância no processo penal, sendo crucial para a construção da convicção do magistrado e um dos elementos chave para a garantia do processo legal, conforme preconizado, por exemplo, na Constituição Federal do Brasil.

Nesse sentido, a formalização matemática do Teorema de Bayes mostra-se — com base em diversos estudos, e como se pretende demonstrar no decorrer deste trabalho — a mais adequada para atingir esse objetivo.

2.1.3. Presunção da não culpabilidade ou presunção de inocência?

Outro princípio constitucional que deve ser entendido a fim de validar o ferramental matemático que será mais adiante apresentado neste trabalho é o **princípio da presunção da inocência**, como ele é impropriamente identificado, ou, de maneira mais apropriada, o **princípio da presunção da não culpabilidade**.

Esses dois princípios são distintos em significação e origem, embora tenham um cerne em comum: não considerar o réu culpado até que se tenha suficiente certeza — formal e material — de sua culpa.

De acordo com Silva (2018b):

O significado de presunção origina-se do latim *praesumptio*, no qual, o verbo é *praesumera*, como bem observam os notáveis, Cretella Júnior e Ulhoa Cintra (1944, p. 896)² ao afirmar que: “significa antecipar, tomar antes ou por primeiro, prever, imaginar previamente”. Portanto, esse termo de origem latina, significa antecipar algo que ainda não aconteceu.

Por outro lado, ainda sob a interpretação dos mesmos autores, inocência, que também provem do latim, *innocentia*, e seu significado está ligado intimamente a práticas religiosas. No campo canônico, a inocência era referida àquele

²CRETELLA, JR., J.; CINTRA, G. D. U. Dicionário Latino-português. São Paulo: COMPANHIA EDITORA NACIONAL-SP, 1944.

que nunca pecou, ou seja, que não infringia as regras divinas, mas com o passar do tempo foi adquirindo uma significação mais filosófica influenciada pelas ideias iluministas que incorporou uma perspectiva mais racional, na qual uma pessoa não poderia ser apontada como culpada sem antes haver uma comprovação que a incriminasse.

A Constituição Federal estabelece, em seu artigo 5º, inc. LVII, que “ninguém será considerado culpado até o trânsito em julgado de sentença penal condenatória” (BRASIL, 1988).

Segundo os instrumentos normativos de Direito Internacional, o acusado é considerado inocente durante toda a ação penal – existe uma presunção, claramente relativa, de que é inocente, e essa presunção de inocência terminará tendo, como marco final, uma sentença condenatória que não precisa transitar em julgado. Segundo a Declaração Universal dos Direitos Humanos (ORGANIZAÇÃO DAS NAÇÕES UNIDAS, 1948):

Art. 11 § 1º – Toda pessoa acusada de um ato delituoso tem o direito de ser presumida inocente até que a sua culpabilidade tenha sido provada de acordo com a lei, em julgamento público no qual lhe tenham sido asseguradas todas as garantias necessárias à sua defesa.

Também segundo a Convenção Americana de Direitos Humanos (ORGANIZAÇÃO DOS ESTADOS AMERICANOS, 1969) em seu Art. 8º § 2º “Toda pessoa acusada de um delito tem direito a que se presuma sua inocência, enquanto não for legalmente comprovada a sua culpa”.

Uma análise não muito profunda permite ver que o que os tratados e convenções internacionais preconizam não é o que está na Constituição Federal, pois, de uma maneira absoluta, utilizando os termos corretos, se vê que a Carta Maior vigente não adotou o princípio da presunção de inocência, adotando, outrossim, o princípio da não culpabilidade.

Assim, pode-se depreender de nosso ordenamento jurídico, a partir dos instru-

mentos de Direito Internacional, que, considerando-se que há uma condenação em segunda instância – ou instância recursal – isso confirmaria a sentença condenatória da primeira instância e já se teria a confirmação da culpa do réu e, com essa confirmação, ele já não poderia ser presumidamente considerado inocente.

Por outro lado, a Constituição Federal afirma que “ninguém será considerado **culpado**(...)” – sem haver presunção de inocência – até o trânsito em julgado da sentença condenatória.

A diferença primordial será no momento do marco final: o marco final dessa presunção de inocência nos tratados internacionais se dá com a sentença condenatória em segunda instância, quando há a confirmação – por parte de um órgão colegiado – daquela condenação imposta a uma pessoa. Já no princípio da presunção da não culpabilidade – adotado na Constituição Brasileira –, o marco final é indubitavelmente o trânsito em julgado da sentença condenatória. Assim, no Brasil, alguém só será considerado culpado – não necessariamente será considerado inocente durante esse intercurso – a partir da sentença condenatória transitada em julgado, ou seja, quando a sentença condenatória não mais puder ser recorrida segundo as normas processuais vigentes.

2.1.3.1. Aspectos linguísticos da *presunção* de inocência e sua relação com o Teorema de Bayes

A avaliação do conjunto probatório por meio do Teorema de Bayes, método que se pretende detalhar nas seções a seguir, consiste, *grosso modo*, em reavaliar constantemente o que se conhece sobre o evento em análise (conhecimentos *a priori*) em função do que vai sendo incorporado ao que se sabe sobre o evento em análise (conhecimentos *a posteriori*). Um dos pressupostos necessários para que se tenha êxito nessa reavaliação é que o conhecimento *a priori* não seja absoluto, ou seja, é **impossível** revisar uma posição que inicialmente consista na certeza absoluta sobre algo.

Nesse ponto, a *presunção* de inocência já foi empregada como argumento crítico ao uso do Teorema de Bayes no meio jurídico quando interpretada sob o axioma “inocente até que se prove culpado” (GIGERENZER et al., 1989, p. 264): uma interpretação simplória dessa frase parece pedir ao julgador que abandone qualquer dúvida sobre a possibilidade de culpa daquele que é julgado e *acredite* tacitamente em sua inocência. Tal abordagem, se de fato condissesse com a correta interpretação lógica desse argumento, impossibilitaria o uso do Teorema de Bayes.

Lançando-se mão da linguística, é possível entender o significado dos verbos epistêmicos³ *presumir* e *acreditar*. Como bem discute Rose (2002, p. 72, tradução minha):

Se estes dois verbos são incompatíveis, então o significado, ou mesmo parte do significado de um deve contradizer o significado, ou mesmo parte do significado do outro. Em outras palavras, *Eu presumo que Y é inocente* contradiz *Eu acredito que Y é inocente*?

Como isto é claramente um problema semântico, é solucionado propriamente pela análise semântica, que é parte da disciplina da Linguística. A ferramenta de análise apropriada é bem conhecida: Metalinguagem Semântica Natural (WIERZBICKA, 1996).

(...)

De acordo com a Metalinguagem Semântica Natural, o significado de *Eu presumo que Y é inocente* é especificado pelos seguintes componentes semânticos, adaptado de A. Wierzbicka (comunicação pessoal):

- **Eu não posso dizer: “Eu sei que Y é inocente”**
- Eu quero pensar que por enquanto: “Eu sei isso”

Por outro lado, *Eu acredito que Y é inocente* envolve os seguintes componentes, adaptado de (GODDARD, 2003):

- A promotoria diz que Y não é inocente
- Quando outros ouvem isso, eles podem pensar que isso é verdade
- Eu não quero pensar que isso é verdade

³Verbos epistêmicos são verbos de pensamento com significados complexos envolvendo primários semânticos tais como *saber*, *verdade*, *dizer*, *pensar* (GODDARD, 2003)

- **Eu penso que Y é inocente**

Rose continua:

Essa partição dos componentes semânticos nos permite ver que não há contradição entre os significados de *presumir* e *acreditar*. Os componentes cruciais são *Eu penso que Y é inocente* (para *acreditar*), e *Eu não posso dizer: “Eu sei que Y é inocente”* (para *presumir*). Estes componentes não são contraditórios, dado que a sentença “Eu penso que é assim mas eu não sei” não implica em contradição. Portanto, pode ser demonstrado por meio da análise NSM que a incompatibilidade das probabilidades *a priori* com a presunção de inocência não é uma crítica válida ao uso legal do teorema de Bayes⁴. (ROSE, 2002, p. 72, tradução minha)

Depreende-se do quanto apresentado que a argumentação acima pode ser diretamente estendida ao preceito legal da *presunção da não culpabilidade* posto que seu cerne recai sobre o mesmo verbo epistêmico (*presumir*) e apenas a presença da dúvida mínima *a priori* contida no componente semântico *Eu não posso dizer: “Eu sei que Y não é culpado”* já é condição suficiente para a validade da aplicação jurídica do Teorema de Bayes no contexto do ordenamento jurídico brasileiro.

⁴If these two verbs are incompatible, then the meaning, or even part of the meaning of one must contradict the meaning, or even part of the meaning of the other. In other words, does *I presume Y is innocent* contradict *I believe Y is innocent*? Since this is clearly a semantic problem, it is properly solved by semantic analysis, which is part of the discipline of linguistics. The appropriate analysis tool is well known: Natural Semantic Metalanguage (Wierzbicka 1996). (...) According to NSM, the meaning of *I presume that Y is innocent* is specified by the following semantic components, adapted from A. Wierzbicka (personal communication): **I can't say: “I know that Y is innocent”**. I want to think for the time being: “I know it”. On the other hand, *I believe that Y is innocent* involves the following components, adapted from Goddard (2001): The prosecution says Y is not innocent. When others hear this, they can think it is true. I don't want to think that it is true. **I think that Y is innocent**. This breakdown of the semantic components allows us to see that there is no contradiction between the meanings of *presume* and *believe*. The crucial components are *I think that Y is innocent* (for *believe*), and *I can't say: “I know that Y is innocent”* (for *presume*). These components are not contradictory, since the statement “I think it is so but I don't know” entails no contradiction. Thus it can be shown by NSM analysis that incompatibility of prior odds with presumption of innocence is not a valid criticism of the legal use of Bayes' theorem. (ROSE, 2002, p. 72)

2.1.4. Aceitabilidade Internacional do Teorema de Bayes na apresentação de resultados periciais

Embora, a esta altura do texto, nada tenha sido ainda apresentado a respeito do que consiste o método matemático do Teorema de Bayes, é interessante apresentar algumas características gerais que fazem com que ele seja o mais adequado para apresentação de resultados periciais.

Saks e Koehler (2005) discutiram a iminente quebra de paradigma que as ciências forenses estavam sofrendo, como um todo, diante da pressão do judiciário e frente ao acúmulo de exemplos de erros em casos reais e da metodologia moderna empregada em testes de DNA. Cabe dizer que os resultados de testes de DNA são necessariamente apresentados em termos de Razão de Verossimilhanças (que decorre do Teorema de Bayes, como será explicado mais adiante na seção 2.3 e subseções).

Morrison (2009) aponta a mudança de comportamento da prática da perícia forense em gradualmente passar a empregar o arcabouço teórico da Razão de Verossimilhanças na apresentação de resultados de avaliações quantitativas. Ele destaca que, apesar de à época do texto, o número de peritos que empregassem a técnica ainda fosse bastante reduzido, era inegável reconhecer que havia uma quebra de paradigma em andamento.

No esteio dessa quebra de paradigma, a Rede Europeia de Institutos de Ciência Forense (European Network of Forensic Science Institutes – ENFSI) recomenda que os resultados periciais avaliativos sejam preferencialmente apresentados em termos de Razão de Verossimilhanças. Perícias avaliativas são o grupo de perícias que requerem uma avaliação por parte do perito a respeito do peso – ou força – da evidência.

O relatório “Diretrizes para Relatórios Avaliativos em Ciência Forense da ENFSI: fortalecendo a avaliação de resultados forenses através da Europa (STEOFRAE)”⁵

⁵ENFSI Guideline for Evaluative Reporting in Forensic Science: Strengthening the Evaluation of Forensic Results across Europe (STEOFRAE)

(WILLIS, 2015) determina que relatórios avaliativos sigam quatro exigências: equilíbrio, lógica, robustez e transparência, conforme segue:

Equilíbrio – os achados devem ser avaliados diante de ao menos duas proposições: usualmente uma que seja baseada na versão dos eventos por uma das partes e uma baseada em uma versão alternativa dos eventos (parte oposta) . Se nenhuma alternativa puder ser formulada, o valor dos achados não pode ser obtido. Nesse caso, peritos forenses devem deixar claro que eles não estão reportando valor sobre os achados.

Lógica – relatórios avaliativos devem referir-se às probabilidades dos achados dadas as proposições e informações relevantes do histórico do evento e não a probabilidade das proposições dados os achados e informações sobre o evento. O relatório não deve conter frases que transponham essa condicional.

Robustez – o relatório deve ser capaz de superar escrutínios e exames cruzados. Deve ser baseado em conhecimentos sólidos e experiências sobre os vestígios e uso de dados. O perito forense estará satisfeito que os resultados das observações e análises sobre as quais inferências e conclusões foram tiradas sejam robustos. Quando houver dados insuficientes, a abordagem por razão de verossimilhanças dará ao perito um arcabouço para uma argumentação estruturada e lógica baseada em sua experiência, desde que ele possa explicar os fundamentos para sua opinião juntamente com seu nível de entendimento a respeito daquele vestígio em particular.

Transparência – as conclusões reportadas devem ser derivadas a partir de processos demonstráveis, tanto no arquivo do caso, quanto no relatório. O relatório deve ser escrito de tal modo que seja adequado para a ampla maioria dos leitores (ou seja, participantes do sistema judiciário). O relatório pode incluir anexos explicando os fundamentos técnicos. (WILLIS, 2015, p. 10-11, tradução minha)⁶

⁶Balance – The findings should be evaluated given at least one pair of propositions: usually one based upon one party's account of the events and one based upon an alternative (opposing party's) account of the events. If no alternative can be formulated, the value of the findings cannot be assessed. In that case, forensic practitioners should state clearly that they are not reporting upon the value of the findings. Logic — Evaluative reports should address the probability of the findings given the propositions and relevant background information and not the probability of the propositions given the findings and background information. The report should not contain statements that are transposing the conditional. Robustness – The reporting should be capable of sustaining scrutiny and cross-examination. It should be based upon sound knowledge and experience of the trace type(s) and the use of data (as defined in the glossary). The forensic practitioner will be satisfied that the results of the observations and analyses upon which inferences and conclusions are drawn are robust. When there are insufficient data, the likelihood ratio approach provides the practitioner with a framework for structured and logical reasoning based on his experience, as long as he can explain the grounds for his opinion together with his degree of understanding of the particular trace type. Transparency – The

O citado relatório da ENFSI vem explicar e unificar a recomendação no âmbito de seus 39 países membros⁷.

Cabe dizer que embora pareça intuitiva a necessidade de se avaliar ao menos duas hipóteses concorrentes, a recomendação mais empregada no Brasil – inclusive ensinada em cursos de comparação facial ministrados pela Polícia Federal do Brasil – ainda é a escala de nove pontos proposta pela Polícia Sueca⁸ (2008, apud ERIKSSON, 2012, tradução minha):

- +4: o resultado corrobora a hipótese com quase toda a certeza;
- +3: o resultado corrobora fortemente a hipótese;
- +2: o resultado corrobora a hipótese;
- +1: o resultado corrobora levemente a hipótese;
- 0: o resultado é inconclusivo;
- -1: o resultado contradiz a hipótese em algum nível;
- -2: o resultado contradiz a hipótese;
- -3: o resultado contradiz fortemente a hipótese;
- -4: o resultado contradiz a hipótese com quase toda a certeza.

Este modelo, por empregar apenas uma hipótese que é avaliada segundo os critérios de conhecimentos do perito forense, acaba por sofrer forte influência subjetiva. Além disso, como emprega apenas uma hipótese – por exemplo, “o vestígio encontrado é do réu” –, tira a isenção do perito forense perante acusação e defesa, posto que apenas uma hipótese é sopesada.

reported conclusions should be derived from a demonstrable process in both the case file and the report (see also 3.11 and 3.12). The report should be written in such a way that it is suitable for a wide audience of readers (i.e., participants in the justice system). It may include supplements explaining the technical background. (WILLIS, 2015, p. 10-11)

⁷Áustria, Armênia, Azerbaijão, Bélgica, Bósnia-Herzegovina, Bulgária, Croácia, Chipre, República Tcheca, Dinamarca, Estônia, Finlândia, França, Geórgia, Alemanha, Grécia, Hungria, Irlanda, Itália, Letônia, Lituânia, República da Macedônia do Norte, Malta, Montenegro, Holanda, Noruega, Polônia, Portugal, Romênia, Rússia, Sérvia, Eslovênia, Eslováquia, Espanha, Suécia, Suíça, Turquia, Ucrânia e Reino Unido (ENFSI, 2020)

⁸Utlåtandeskalan (2008) <<http://www.skl.polisen.se/Global/www%20och%20Intrapolis/Informationsmaterial/SKL/Utlatandeskalan.pdf>>

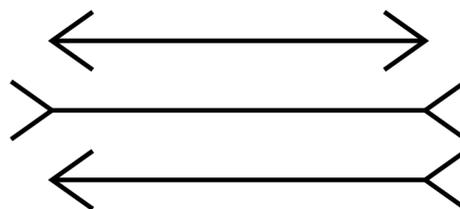
Mais adiante, na seção 2.3.5.3, essa escala será comparada com a apresentação dos resultados periciais por meio da Razão de Verossimilhanças.

2.2. Vieses cognitivos: uma visão geral

Vieses cognitivos são, em termos bastante simplificados, situações em que a cognição humana produz sistematicamente representações ou conclusões distorcidas em comparação a algum aspecto objetivo da realidade.

É comum que a primeira percepção a respeito dos vieses cognitivos, quando se trava contato com sua existência, é a de que se tratem de *falhas* de raciocínio. Um exemplo visual bastante conhecido de como o processo cognitivo pode levar a conclusões errôneas é a “ilusão de Müller-Lyer” (figura 2.1):

Figura 2.1. – “Ilusão de Müller-Lyer”.



Exemplo da “ilusão de Müller-Lyer”. Qual flecha é maior? Qual o ponto central de cada flecha? Fonte: (MÜLLER-LYER ILLUSION, 2021).

Nesse exemplo, embora sejamos levados a crer que as linhas possuem comprimentos distintos, todas elas têm igual comprimento. A “ilusão de Müller-Lyer” é uma ilusão visual que serve como exemplo motivador para o estudo dos processos cognitivos.

Os vieses cognitivos compõem campo de estudo da psicologia evolucionista (HASELTON; NETTLE; ANDREWS, 2015) que busca interpretar a presença dos diversos vieses como sendo resultados dos desafios evolutivos que a espécie humana enfrentou. O presente texto não pretende explorar as bases evolutivas dos vieses

cognitivos, mas, tão somente apresentar alguns daqueles que têm conexão com a proposta aqui apresentada e situá-los como uma característica inerente e presente nos processos decisórios humanos e que devem ser compreendidos e confrontados a fim de se buscar estratégias para contorná-los.

Diversos ramos do conhecimento humano, reconhecem o papel – prejudicial em alguns contextos – dos vieses cognitivos nas decisões de seus profissionais.

Na medicina, vieses cognitivos podem levar a erros médicos. Por exemplo, quando um médico deixa de diagnosticar uma determinada doença em um paciente, especialmente alguma com baixa prevalência populacional, e é confrontado por exames ou pela própria evolução do quadro clínico daquele paciente, esse médico tende a ter aquela doença como sua primeira suspeita quando sintomas semelhantes se apresentarem em outros pacientes. Esse médico estará sujeito a solicitar exames que, de outra forma, não teriam indicações objetivas e, portanto, ele passará a estar sujeito a interpretações errôneas. Trata-se do denominado *viés da disponibilidade*⁹: respostas e soluções mais recentes e prontamente disponíveis são preferencialmente favorecidas por sua facilidade de serem recordadas e incorreta percepção de importância (O’SULLIVAN; SCHOFIELD, 2018). O’Sullivan e Schofield (2018) trazem uma lista não exaustiva dos vieses cognitivos presentes em clínica médica. Bhatti (2018) propõe estratégias para desenvolver *habilidades metacognitivas* em estudantes de medicina para que, já durante suas formações, sejam expostos a técnicas de desenviesamento¹⁰ cognitivo.

Outro viés que merece destaque é o *viés da retrospectiva*¹¹ também conhecido como *efeito “eu sabia disso o tempo todo”* (FISCHHOFF; BEYTH, 1975) (FISCHHOFF, 1975). Em linhas gerais, é a tendência a exagerar retrospectivamente a qualidade de uma previsão sobre o resultado de um evento. Chen et al. (2019) solicitaram que os participantes do estudo fizessem predições sobre o resultado de eleições

⁹ *availability bias.*

¹⁰ *debiasing.*

¹¹ *hindsight bias.*

municipais do ano de 2014 para seis cidades de Taiwan e, após as eleições, eles foram novamente consultados para lembrar e avaliar a qualidade de suas predições. Os resultados mostraram que 88 % dos participantes apresentaram viés da retrospectiva ao lembrarem que suas predições foram mais próximas dos resultados reais do que realmente foram. Hoffrage, Hertwig e Gigerenzer (2000) propuseram que o viés da retrospectiva seria um subproduto de processo adaptativo em que o sujeito atualiza seu conhecimento após receber *feedback*.

No Direito e nas Ciências Forenses, o papel dos vieses cognitivos nas decisões de jurados, juízes e na produção e apresentação de exames periciais também é reconhecido e estudado.

Harley (2007) estudou o viés da retrospectiva em decisões legais em julgamentos a respeito de responsabilidade legal por erros médicos e fez uma revisão dos estudos que buscavam reduzir ou eliminar esse viés dos tribunais. Salerno e Diamond (2010) fizeram uma revisão sobre os trabalhos que relacionam processos cognitivos durante a deliberação do júri. Bilotta et al. (2019) produziram uma revisão a respeito dos *vieses sutis* presentes em formas contemporâneas de preconceitos e estereótipos que, muitas vezes inconscientemente, afetam procedimentos do sistema legal.

Kukucka et al. (2017) pesquisaram junto a 403 peritos experientes, em 21 países, a respeito de suas crenças quanto à natureza e escopo dos vieses cognitivos. Os autores concluíram que:

Muitos peritos acreditam que são imunes aos vieses ou que podem reduzi-los através de mera força de vontade, e pouco menos da metade deles apoiam testes cegos. Além disso, muitos peritos apresentaram o *viés do ponto cego* (PRONIN; LIN; ROSS, 2002), reconhecendo vieses em outros campos dos conhecimento mas não em seus próprios campos, e em outros peritos, mas não neles mesmos. Estes achados reforçam a necessidade de reformas procedurais que escondam dos peritos forenses informações potencialmente enviesantes, como é comum em outros ramos da ciência.¹² (KUKUCKA et al.,

¹²Most examiners believed they are immune to bias or can reduce bias through mere willpower, and fewer than half supported blind testing. Furthermore, many examiners showed a bias blind spot (PRONIN;

2017, tradução minha)

Zapf et al. (2018) estudaram profissionais em saúde mental que produzem avaliações forenses no âmbito judicial. Esses autores assim resumiram seus achados:

A maioria dos avaliadores expressou preocupação a respeito dos vieses cognitivos *mas* manteve uma visão incorreta de que mera força de vontade pode reduzir vieses. Evidência também foi encontrada para um *viés do ponto cego* (PRONIN; LIN; ROSS, 2002), com mais avaliadores reconhecendo vieses nos julgamentos feitos por seus pares do que em seus próprios. Avaliadores que receberam treinamento sobre vieses foram mais propensos a reconhecer vieses cognitivos como causa para preocupação em avaliações forenses bem como em seus próprios julgamentos.¹³ (ZAPF et al., 2018, tradução minha)

Fraser (2019) chama a atenção quanto a uma tarefa usualmente considerada de pouca ou nenhuma complexidade na perícia forense: a transcrição de áudios. Ela demonstrou, empregando exemplo real de um julgamento por assassinato, que a transcrição feita por policiais foi imprecisa posto que eles estavam altamente influenciados pelas informações do caso em si: efeito de pré-ativação¹⁴. A autora concluiu que áudios de baixa qualidade somente deveriam ser empregados como evidência se acompanhados por uma transcrição confiável e independente¹⁵.

LIN; ROSS, 2002), acknowledging bias in other domains but not their own, and in other examiners but not themselves. These findings underscore the necessity of procedural reforms that blind forensic examiners to potentially biasing information, as is commonplace in other branches of science.

¹³Most evaluators expressed concern over cognitive bias but held an incorrect view that mere willpower can reduce bias. Evidence was also found for a bias blind spot (PRONIN; LIN; ROSS, 2002), with more evaluators acknowledging bias in their peers' judgments than in their own. Evaluators who had received training about bias were more likely to acknowledge cognitive bias as a cause for concern, whereas evaluators with more experience were less likely to acknowledge cognitive bias as a cause for concern in forensic evaluation as well as in their own judgments. Training efforts should highlight the bias blind spot and the fallibility of introspection or conscious effort as a means of reducing bias. In addition, policies and procedural guidance should be developed in regard to best cognitive practices in forensic evaluations.

¹⁴*priming effect*.

¹⁵Cabe apontar que, no Estado de São Paulo, em consonância às boas práticas apontadas por Fraser, a Resolução nº 142/14 da Secretaria de Segurança Pública (SÃO PAULO, 2014) determina que:

Artigo 1º – Compete ao Delegado de Polícia que conduz o procedimento de interceptação telefônica, informática ou telemática, a transcrição, nos termos da legislação vigente das comunicações gravadas em mídia que não exijam conhecimento técnico-científico.

Parágrafo 1º – A transcrição será realizada por policial civil subordinado designado pela autoridade policial e, apenas excepcionalmente, será requisitada ao Instituto de Criminalística.

No artigo “Conditions for Intuitive Expertise”¹⁶, Kahneman e Klein (2009) buscam delinear as condições limites que separam a *verdadeira habilidade intuitiva* das *impressões enviesadas e com excesso de confiança*. Dentre as conclusões desse trabalho, cabem destaque:

- Julgadores habilidosos muitas vezes desconhecem as pistas que os guiam, e indivíduos cujas intuições não são hábeis têm ainda menos chance de saberem a fonte de seus julgamentos.¹⁷
- Diz-se que verdadeiros especialistas sabem quando eles não sabem. Entretanto, não-especialistas (independente de eles pensarem ou não que o são) certamente não sabem quando eles não sabem. Confiança subjetiva é portanto uma indicação não confiável de validade de julgamentos e decisões intuitivas. (KAHNEMAN; KLEIN, 2009, tradução minha)¹⁸

Ward et al. (2018) produziram uma revisão de 140 publicações que explicitamente relacionam a necessidade da *habilidade adaptativa* como condição *sine qua non* para a perícia (perícia em seu sentido mais geral como domínio de um determinado campo do saber – *expertise*). Esse texto objetivava dar base para guiar políticas de treinamento e educação na área de desenvolvimento de habilidade cognitivas complexas que são base para natureza adaptativa da perícia e *performance* adaptativa especializada. Os autores concluem:

Os princípios de treinamento que listamos advogam pelo uso de oportunidades de aprendizado em casos desafiadores e não-rotineiros, que sejam baseados nas experiências vividas por peritos, e que sejam focadas em problemas, colaborativas e comandadas por facilitadores especialistas. Esses princípios encorajam alunos a se engajarem no uso das habilidades do pensamento

Artigo 2º – As transcrições ou análises das comunicações gravadas em mídia que exijam conhecimentos técnico-científicos deverão ser requisitadas ao Instituto de Criminalística para elaboração de laudo pericial nos termos do Código de Processo Penal.

¹⁶Condições para a Expertise Intuitiva.

¹⁷Skilled judges are often unaware of the cues that guide them, and individuals whose intuitions are not skilled are even less likely to know where their judgments come from.

¹⁸True experts, it is said, know when they don't know. However, nonexperts (whether or not they think they are) certainly do not know when they don't know. Subjective confidence is therefore an unreliable indication of the validity of intuitive judgments and decisions.

crítico e a fazerem o aprendizado sensível ao contexto. Os princípios advogam pela customização do aprendizado através da proficiência escalonada, e a ampliar as oportunidades de aprendizado cara-a-cara explorando e mesclando com tecnologias apropriadas, tais como sistemas *online*, simulações, realidade virtual, e tecnologia de jogos educativos quando necessários. Há forte evidência que treinamentos que fazem uso dos princípios acima são efetivos na aceleração do desenvolvimento de proficiência, e há evidência sugestiva que estes princípios podem promover o desenvolvimento desses tipos de habilidades e características (i.e., entendimento, ação flexível e metacognição) que são consistentes com a noção de habilidade adaptativa. (WARD et al., 2018, tradução minha)¹⁹

O juiz federal Eduardo Costa, em sua tese de doutoramento (COSTA, 2016), aborda o papel dos vieses cognitivos como dificultadores da promoção da imparcialidade judicial. Ele apresenta aspectos psicológicos e comportamentais sobre alguns dos vieses cognitivos que, na sua visão, permeiam de modo prevalente os processos decisórios de juízes. Também são propostas abordagens para neutralização-eliminação-mitigação desses vieses. Costa (2017) explica que “o direito probatório é o signo supremo da perene precariedade dos julgamentos humanos” posto que “a busca pela verdade em estado de pureza bruta está condenada a uma *quintúplice indigência fundamental*”: i) *finitude compreensiva do juiz*, ii) *modo privativo da verdade no processo*, iii) *modo de ser corruptível do jogo processual*, iv) *opacidade da facticidade* e v) *occlusão temporal da descoberta*. Destes, relaciona-se aos vieses cognitivos humanos a *finitude compreensiva do juiz*, assim definida como:

i) a *finitude compreensiva do juiz* (que não é onisciente nem onipotente, sendo o direito probatório o clímax representativo da imanência, da historicidade e do

¹⁹The training principles we listed advocate the use of learning opportunities on challenging and non-routine cases, that are based on the expert’s lived experiences, and that are problem-focused, collaborative, and led by skilled facilitators. These principles encourage learners to engage in the use of critical-thinking skills and to make learning context-sensitive. The principles advocate for the customization of learning through appropriate proficiency scaling, and to enhance face-to-face learning opportunities by exploiting and blending with appropriate technology, such as online systems, simulation, virtual reality, and educational gaming technology as necessary. There is strong evidence that training that leverages the above principles is effective in accelerating the development of proficiency, and there is suggestive evidence that these principles can promote development of the kinds of skills and characteristics (i.e., understanding, flexible action, and meta-cognition) that are consistent with the notion of adaptive skill..

limite indisponíveis em que mergulhado o juiz “fatural”, a experiência normativa privilegiada que nos abre para a atmosfera existencial da finitude judicial) (COSTA, 2017).

Em sua dissertação de mestrado, Silva (2018a) relaciona a necessidade do enfrentamento ativo aos vieses cognitivos no meio jurídico como passo fundamental ao chamado Processo Democrático. As soluções propostas pelo autor, que podem parecer até certo ponto simplistas, posto que assemelham-se a atos de *força de vontade* das partes, envolvem:

(...) uma efetiva campanha de demonstração de existência e aplicação dos vieses de cognição aos magistrados, aliada ao correto dimensionamento da fase preparatória do processo – por meio do contraditório de influência e da fundamentação racional das decisões – e da colegialidade recursal – por meio do incentivo ao dissenso – são capazes de proporcionar o necessário controle da jurisdição por meio do *debiasing*, levando a decisões menos afetadas pelos vieses de cognição e pela subjetividade do julgador (...).

Inobstante, concluiu-se também que a atuação ativa dos advogados, valendo-se sempre que possível das prerrogativas do princípio da oralidade, é fundamental para a almejada *accountability*, seja por meio de despachos com os magistrados, apresentação de memoriais, ou realização de sustentação oral, no sentido de se criar uma identificação da demanda e chamar a atenção para os contornos da controvérsia e para detalhes probatórios, estabelecendo-se uma incansável tentativa de quebra do enviesamento a que possam estar acometidos os julgadores.

Por fim, o que se observa é que cabe aos operadores do direito – tanto aos advogados, quanto aos juízes, pensadores e demais – buscar uma atuação contrafática, de modo a realizar a identificação e a quebra dos enviesamentos cognitivos.

Diante de todo o exposto, nota-se que o impacto dos vieses cognitivos, inerentemente presentes na cognição, deve ser levado em conta em todas as atividades humanas. No âmbito jurídico, deve ser algo amplamente entendido e avaliado tanto por quem realiza os exames periciais, quanto por quem julga. O sujeito que julga o peso de uma prova ou que julga o peso relativo de um conjunto probatório deve conscientizar-se

de suas conclusões intuitivas e cercar-se de ferramentas e técnicas que o auxiliem nessa tarefa.

2.2.1. “Paradoxo de Monty Hall” e a ilusão do controle

A ilusão do controle é um tipo de viés cognitivo que pode assim ser definida: “(...) uma expectativa de uma probabilidade de sucesso pessoal inapropriadamente mais alto do que a probabilidade objetiva pode garantir”²⁰ (LANGER, 1975, tradução minha). A fim de motivar a percepção do papel desse viés cognitivo nos processos decisórios, faz-se necessário introduzir o “Dilema de Monty Hall” (MHD²¹) (ou “Paradoxo de Monty Hall” ou ainda “Problema de Monty Hall”) que é um enigma probabilístico notadamente conhecido por induzir pessoas a decisões incorretas – ou subótimas, na terminologia usual – e que ficou bastante conhecido por fazer parte do programa de televisão “*Let’s Make a Deal*” apresentado na década de 1970 nos Estados Unidos. Seu enunciado é simples:

1. O apresentador do programa mostra três portas a um participante;
2. Atrás de apenas uma das portas há um prêmio;
3. O participante, em busca do prêmio, escolhe uma das portas, que é mantida fechada;
4. O apresentador, que sabe em qual porta encontra-se o prêmio, abre uma das duas portas restantes e mostra seu interior vazio;
5. Ao participante é dada uma nova oportunidade de escolha entre **permanecer com a porta que escolheu inicialmente** ou **alterar sua escolha para a outra porta restante**.

Qual estratégia é a mais adequada para maximizar a chance de que o partici-

²⁰(...) an expectancy of a personal success probability inappropriately higher than the objective probability would warrant.

²¹Do inglês, *Monty Hall Dilemma*.

pante consiga o prêmio? Afinal, agora restam apenas duas portas, logo, a chance do prêmio estar em uma das portas é de 50 %! Será?

Há muitas formas de solucionar esse problema e concluir que a melhor estratégia, embora contraintuitiva, seja sempre alterar a escolha para a outra porta restante para que a chance de vitória aumente de um terço para dois terços. Mais a frente, neste trabalho, será apresentada a abordagem desse paradoxo por meio do Teorema de Bayes – e a inclusão matemática de um fator que facilmente passa despercebido. Argumentativamente, uma explicação bastante didática é a que segue:

“A probabilidade que um participante selecione aleatoriamente a porta com o prêmio em sua primeira tentativa é de uma em três, porque há um prêmio em três locais possíveis. Se a seleção inicial for a porta vencedora, então depois que o apresentador revela uma porta sem prêmio, conclui-se que a segunda porta sem prêmio seja a porta restante não selecionada. Portanto, em um terço de todas as tentativas, permanecer com a porta escolhida vence e trocar perde. Em contraste, a probabilidade que um participante não escolha a porta com o prêmio em sua primeira tentativa é de duas em três. Se a porta escolhida pelo participante não contém prêmio, e outra porta sem prêmio é revelada, então a porta restante deve conter o prêmio. Portanto, em dois terços das vezes, trocar de porta vence e permanecer perde.²²” (HERBRANSON, 2012, tradução e adaptação nossa).

Granberg e Brown (1995) observaram que, num primeiro contato com o problema, apenas 12 % dos indivíduos optaram pela troca de portas e, após sucessivas tentativas, 55 % dos indivíduos passaram a escolher pela troca de portas. As pessoas sentiam-se frustradas e bravas se trocavam de porta e perdiam. Levantou-se a hipótese de que esse comportamento fosse relacionado ao viés cognitivo da ilusão de controle.

²²The probability that a participant will randomly select the door with the prize on the first guess is one in three, because there is one prize and three possible locations. If the initial selection is the winning door, then after the host reveals one goat, it follows that the second goat must be behind the remaining unselected door. Thus, on one-third of all trials, staying wins and switching loses. In contrast, the probability that a contestant will not choose the door with the prize on the first guess is two in three. If the contestant's chosen door conceals one goat, and the other goat has been revealed, then the remaining door must conceal the prize. Thus, two-thirds of the time, switching wins and staying loses. (HERBRANSON, 2012)

2.2.1.1. Dissonância cognitiva

Em (GILOVICH; MEDVEC; CHEN, 1995), uma versão adaptada do MHD foi empregada a fim de estudar os efeitos cognitivos dos erros por omissão (permanecer com uma porta e descobrir que não há prêmio atrás dela) ou por comprometimento – ou ação – (optar por trocar de porta e descobrir que o prêmio estava atrás da porta inicialmente escolhida) sob a ótica da psicologia do arrependimento e da dissonância cognitiva. Concluiu-se que, em eventos cujos resultados se desenrolam em curto espaço de tempo, os indivíduos tendem a reduzir a dissonância cognitiva mais para seus erros de comprometimento do que para erros de omissão. Ou seja, infere-se que seja mais simples aceitar e lidar com a perda originada por uma ação do que por uma omissão.

2.2.1.2. Efeito Intercultural

Granberg (1999) estudou o efeito intercultural em estudantes no Brasil, na China, na Suécia e nos Estados Unidos quando apresentados a tentativas únicas do Dilema de Monty Hall e do “Dilema da Roleta Russa” (RRD²³), que é o problema simetricamente oposto ao MHD em que duas portas guardam prêmios e uma porta não tem nada. Como já explicado, a estratégia ótima (solução racional) para o MHD é trocar de porta em todas as tentativas. Já no RRD, por uma argumentação análoga, a estratégia ótima consiste em sempre permanecer com a porta escolhida inicialmente. Os resultados (ver tab. 2.1) mostraram que, independentemente da versão do problema, humanos tendem ao comportamento de valorizar sua escolha inicial, corroborando os achados anteriormente citados de Granberg e Brown (1995). A comparação entre dados de culturas distintas, não resultou em evidências de comportamentos ou tendências significativamente diferentes entre essas quatro culturas.

²³Do inglês, Russian Roulette Dilemma.

Tabela 2.1. – Percentuais de escolha pela permanência na porta escolhida inicialmente no MHD e no RRD em Brasil, China, Suécia e EUA.

	Brasil	China	Suécia	EUA
MHD	87	79	83	84
RRD	78	81	87	85
MHD (homens)	80 (49)	75 (48)	83 (76)	72 (40)
RRD (homens)	78 (51)	75 (51)	92 (84)	84 (43)
MHD (mulheres)	94 (51)	83 (54)	84 (98)	93 (56)
RRD (mulheres)	78 (49)	89 (46)	83 (88)	86 (51)

Os números entre parênteses são a quantidade de participantes em cada condição. Fonte: (GRANBERG, 1999, tradução minha).

2.2.1.3. Comparação entre espécies: humanos x pombos

Herbranson e Schroeder (2010) fizeram um experimento em que compararam o comportamento de seres humanos e pombos diante de repetições consecutivas do Dilema de Monty Hall. Enquanto uma tentativa única pode indicar a capacidade de compreensão inicial do problema, a repetição consecutiva busca identificar a habilidade de aprendizado de cada grupo. Nesse artigo de título provocativo – *Are Birds Smarter Than Mathematicians? Pigeons (Columba livia) Perform Optimally on a Version of the Monty Hall Dilemma*²⁴ –, os autores descreveram que empregaram um aparato que simulava o dilema de Monty Hall a fim de treinar pombos. Eles foram treinados e completaram até 100 repetições do teste por dia, durante aproximadamente 30 dias. Os participantes humanos completaram 200 testes no total. Os humanos não receberam qualquer orientação a respeito dos detalhes do problema sendo apenas orientados a tentarem maximizar o número de vitórias.

Ao discutir os resultados (ver figura 2.2), Herbranson (2012) conclui que, por

²⁴Pombos são mais inteligentes que matemáticos? Pombos têm resultados ótimos numa versão do Dilema de Monty Hall.

alguma razão, humanos não são adequados para resolverem o Dilema de Monty Hall:

(...) Pombos começam com uma tendência a não mudarem, mas finalmente decidem por uma estratégia de troca em virtualmente todas as tentativas. Participantes humanos rapidamente desenvolvem uma tendência a trocar em cerca de dois terços das tentativas. Os humanos completaram menos testes do que os pombos, mas animais geralmente requerem mais treinamento em estudos comparativos porque não podem receber instruções verbais e devem aprender procedimentos por tentativa e erro. Mais importante, as tendências de resposta dos humanos não variam significativamente em torno das 50 tentativas finais, indicando que eles atingiram um platô estável²⁵ (HERBRANSON, 2012, tradução minha).

E ainda:

(...) As razões para a diferença na performance entre espécies no Dilema de Monty Hall são neste momento incertas, mas elas podem originar-se em estratégias para busca de alimentos em ambientes desiguais (CHARNOV, 1976). Um coletor ótimo permanecerá em um caminho ou trocará para outro baseando-se em fatores tais como a riqueza do ambiente e a energia requerida para se movimentar. Se o Dilema de Monty Hall se baseia nos mesmos processos cognitivos usados em busca de alimentos, então animais de diferentes nichos ecológicos podem ser diferentemente inclinados a adotar rapidamente uma estratégia de troca.²⁶ (HERBRANSON, 2012, tradução minha)

Numa abordagem complementar aos elementos trazidos até aqui, propõe-se o uso do Teorema de Bayes a fim de mitigar os vieses cognitivos na avaliação de provas técnicas periciais.

²⁵Pigeons began with a tendency to stay, but they eventually settled on a strategy of switching on virtually all trials. Human participants quickly developed a tendency to switch on about two-thirds of trials. The humans completed fewer trials than the pigeons did, but animals generally require more training in comparative studies because they cannot be given verbal instructions and must learn procedures through trial and error. More important, the response tendencies of humans did not vary significantly over the final 50 trials, indicating that they had reached a stable plateau. (HERBRANSON, 2012, p. 298)

²⁶The reasons for the interspecies difference in performance on the Monty Hall Dilemma are at this point unclear, but they might originate in strategies for optimal foraging in patchy environments (CHARNOV, 1976). An optimal forager will stay at one patch or switch to another on the basis of factors such as the richness of the environment and the energy required to move about. If the Monty Hall Dilemma draws on the same cognitive processes used in foraging, then animals from different ecological niches might be differently inclined to quickly adopt a switching strategy. (HERBRANSON, 2012)

Figura 2.2. – Comparação entre pombos e humanos diante do Dilema de Monty Hall.

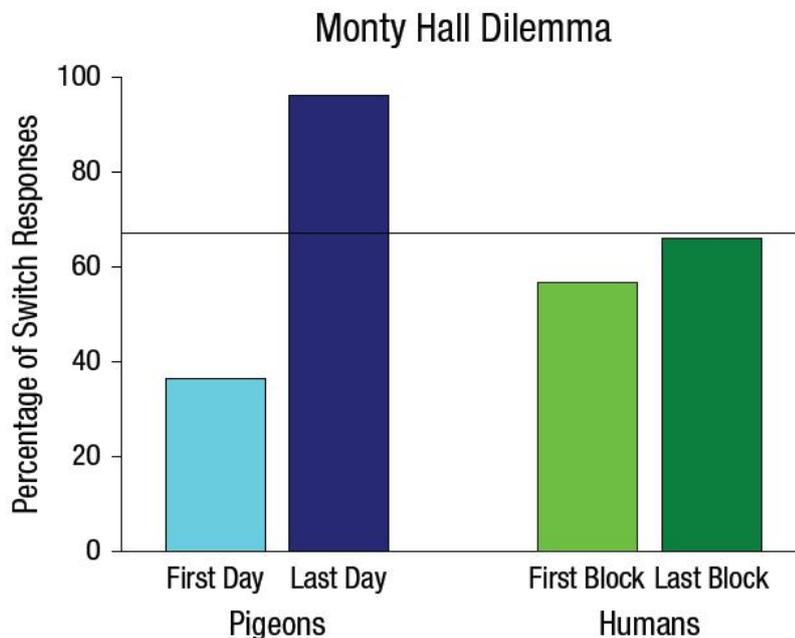


Fig. 2. Percentage of switch responses by pigeons and humans at the beginning and end of the experiment by Herbranson and Schroeder (2010). The line in the graph is at 67%, the probability of winning with a switch response. The optimal strategy is to switch on all trials.

Percentual de respostas de troca de porta por pombos e humanos no início e no final do experimento por Herbranson e Schroeder (2010). A linha no gráfico demarca 67 %, a probabilidade de vencer com a resposta de troca. A estratégia ótima é trocar em todas as tentativas. Fonte: (HERBRANSON, 2012, p. 298, tradução minha).

2.3. Teorema de Bayes

Existem duas principais abordagens estatísticas: descritiva e inferencial. A estatística descritiva restringe-se à exploração dos dados por meios de tabelas, gráficos e pela obtenção de indicadores matemáticos, tais como médias e desvio padrão. Já a abordagem da estatística inferencial procura extrair conclusões – inferências – por meio da análise dos dados.

A estatística inferencial divide-se em dois ramos: inferência frequentista – ou clássica – e inferência bayesiana. Na inferência clássica, as previsões devem se restringir à análise dos dados disponíveis. Exemplo bastante conhecido são as pesquisas eleitorais, em que se pretende inferir o comportamento de toda a população por meio

de um extrato das escolhas de uma parcela representativa dessa população. Na inferência bayesiana, é possível incluir informações subjetivas sobre os dados disponíveis – conhecimentos *a priori* – a fim de se ampliar o escopo dessa inferência sobre os dados.

O Teorema de Bayes é uma lei para descrição de probabilidades em um evento posterior (*a posteriori*) com base em conhecimentos anteriores (*a priori*) que podem ou não estar diretamente relacionados a algum evento principal. Trata-se de uma formulação matemática que se propõe a sistematizar a inclusão de informações subjetivas ou de fontes relacionadas – ou mesmo alheias – aos dados auxiliando o processo decisório humano de modo a evitar as idiosincrasias advindas dos mais variados vieses cognitivos.

A pesquisa de conteúdo a respeito desse tema mostrou-se, em grande parte, com linguagem e abordagens inacessíveis ao público geral, seja por tratarem o assunto com foco no formalismo matemático, que dificulta a compreensão prática da técnica, seja pelo excesso de simplificações do uso técnico, que deixam a desejar no entendimento mínimo do formalismo matemático necessário à generalização do método.

Diante da abordagem interdisciplinar empregada no presente texto, o Teorema de Bayes será apresentado em termos simplificados, com exemplos práticos, usuais e de resultados inicialmente contraintuitivos que são muitas vezes desconhecidos aos públicos leigo e mesmo técnico, mas sem deixar de detalhar e formalizar as definições empregadas em sua dedução.

Para uma abordagem introdutória sobre o Teorema de Bayes sugerem-se os *podcasts* (FUJIOKA; SOUZA, 2017) e (MALTA et al., 2020).

Para maior aprofundamento sobre esse importante tema da teoria das probabilidades e da estatística, sugere-se (WASSERMAN, 2003), (DEKKING et al., 2005) e (HOIJTINK; KLUGKIST; BOELEN, 2008). A referência (TARONI et al., 2010) traz uma abordagem focada em ciências forenses.

2.3.1. Conceitos fundamentais

A explicação adotada nesta seção foi inspirada na referência (NASCIMENTO, 2017).

A notação que será adotada neste texto é a que segue: $P(A)$ é a probabilidade de ocorrência de um evento A em um conjunto específico de valores possíveis; $P(A|B)$ é a probabilidade condicional de ocorrência de um evento A dada a ocorrência do evento B .

Para derivar a equação do Teorema de Bayes faz-se necessário lançar mão de algumas definições do campo da probabilidade, a saber: 1) Independência, 2) Probabilidade Conjunta, 3) Permutabilidade e 4) Regra da Probabilidade Total (PPT).

O problema norteador para o desenvolvimento dessas definições será este: imagine-se uma caixa contendo 2 bolas brancas e 3 bolas pretas empregadas em um experimento de retirada ao acaso, sem reposição, e registro da cor da bola. Pode-se definir dois eventos distintos:

- A : a bola retirada é da cor branca;
- B : a bola retirada é da cor preta;

2.3.1.1. Independência

A independência entre eventos X_i , com $i = 1, 2, \dots, N$, ocorre quando a probabilidade de um dado evento X_i não permite inferir nenhuma conclusão a respeito de um evento X_j distinto ($i \neq j$).

Antes que seja retirada alguma bola, as probabilidades iniciais são: $P(A) = \frac{2}{5}$ e $P(B) = \frac{3}{5}$.

Nota-se que essas probabilidades não se manterão constantes após a primeira retirada posto que a população de bolas será alterada. Pode-se concluir que, caso a primeira bola retirada seja branca (evento *a priori* A_1), então, as probabilidades da

segunda retirada (eventos A_2 e B) serão: $P(A_2|A_1) = \frac{1}{4}$ e $P(B|A_1) = \frac{3}{4}$.

Mas, caso a primeira bola retirada seja preta (evento *a priori* B_1), então, as probabilidades da segunda retirada (eventos A e B_2) serão: $P(A|B_1) = \frac{1}{2}$ e $P(B_2|B_1) = \frac{1}{2}$.

Na segunda retirada, as probabilidades são condicionadas por um evento anterior, portanto, não são independentes.

2.3.1.2. Probabilidade Conjunta

A probabilidade conjunta de dois eventos A e B é expressa pela regra do produto:

$$P(B \cap A) = P(B|A)P(A) \quad (2.1)$$

Empregando ainda o exemplo da retirada de bolas, a probabilidade conjunta de que a primeira retirada seja de uma bola branca (evento A) e a segunda retirada seja de uma bola preta (evento B) é igual a: $P(B \cap A) = P(B|A)P(A) = \frac{3}{4} \cdot \frac{2}{5} = \frac{3}{10}$.

2.3.1.3. Permutabilidade

Permutabilidade é a propriedade da alteração no ordenamento de realizações em uma série de eventos sem que a probabilidade conjunta seja alterada. É expressa pela fórmula:

$$P(A|B)P(B) = P(A \cap B) = P(B \cap A) = P(B|A)P(A) \quad (2.2)$$

Como já explicado, a probabilidade conjunta de que a primeira retirada seja de uma bola branca (evento A) e a segunda retirada seja de uma bola preta (evento B) é igual a $P(B \cap A) = \frac{3}{10}$.

Invertendo-se essa ordem, é possível calcular a probabilidade conjunta de que a primeira retirada seja de uma bola preta (evento B) e a segunda retirada seja de uma

bola branca (evento A) como igual a $P(A \cap B) = P(A|B)P(B) = \frac{1}{2} \cdot \frac{3}{5} = \frac{3}{10}$.

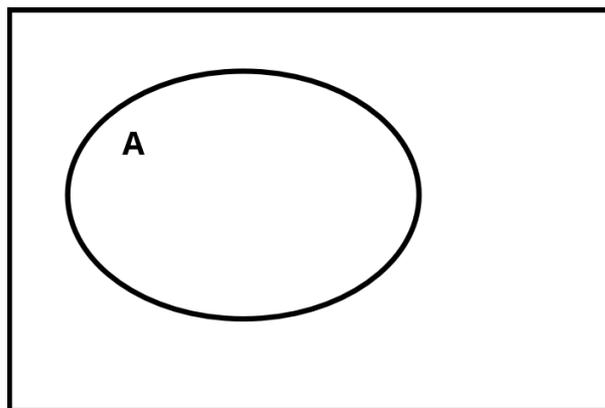
Portanto, como $P(A \cap B) = P(B \cap A)$, trata-se de um experimento de probabilidade permutável.

2.3.1.4. Regra da Probabilidade Total

Expressa a probabilidade total de um resultado por meio de vários eventos disjuntos (mutuamente exclusivos).

Suponha o problema de encontrar o valor da probabilidade do evento A em um espaço de valores possíveis Ω (figura 2.3).

Figura 2.3. – Probabilidade do evento A em um espaço amostral Ω .



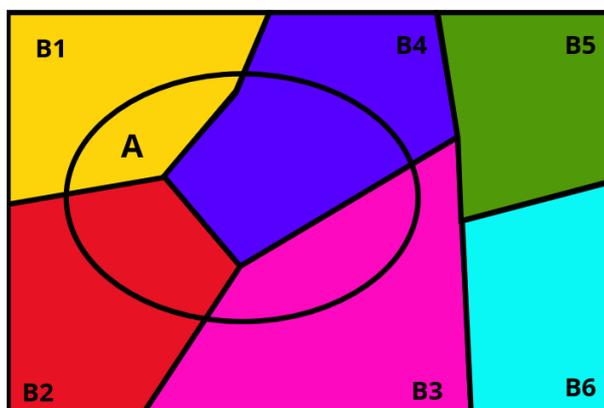
Fonte: (NASCIMENTO, 2017).

Imagine que o espaço Ω pode ser dividido entre as probabilidades de eventos B_i sem intersecções entre si, ou seja, eventos mutuamente exclusivos (figura 2.4).

Mesmo desconhecendo-se a natureza da probabilidade do evento A , é possível escrever esse valor em função dos blocos formados pelas intersecções entre o evento A e cada partição dos eventos B_i (figuras 2.5 e 2.6).

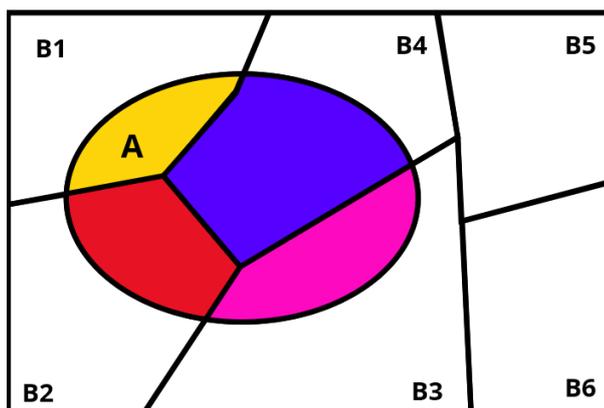
Generalizando-se esse resultado, nos espaços amostrais Ω formados pela união de partes B_i disjuntas, e cuja união seja o espaço amostral, a probabilidade de qualquer evento A de Ω é:

Figura 2.4. – Espaço amostral Ω dividido em probabilidades de eventos B_i , sem intersecções entre si.



Fonte: (NASCIMENTO, 2017).

Figura 2.5. – Probabilidade do evento A em função de partições dos eventos B_i .



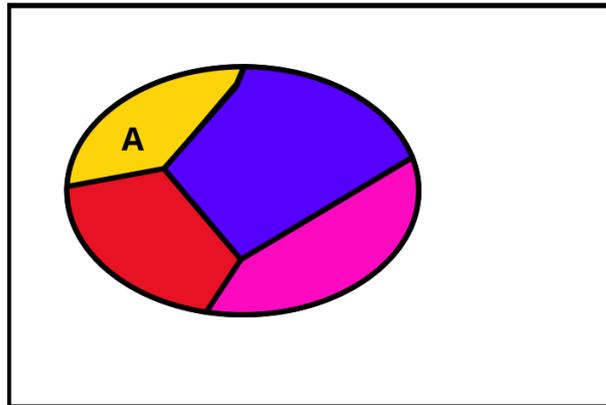
Fonte: (NASCIMENTO, 2017).

$$\begin{aligned}
 P(A) &= P(A \cap B_1) + P(A \cap B_2) + \dots + P(A \cap B_N) \\
 &= P(A|B_1)P(B_1) + P(A|B_2)P(B_2) + \dots + P(A|B_N)P(B_N) \\
 &= \sum_{i=1}^N P(A|B_i)P(B_i)
 \end{aligned} \tag{2.3}$$

2.3.2. Formulação do Teorema de Bayes

O Teorema de Bayes pode ser então derivado a partir da definição de probabilidade conjunta (seção 2.3.1.2):

Figura 2.6. – Probabilidade do evento A em função de partições dos eventos B_i .



Fonte: (NASCIMENTO, 2017).

$$P(A|B) = \frac{P(A \cap B)}{P(B)}, \text{ se } P(B) \neq 0 \quad (2.4)$$

$$P(B|A) = \frac{P(B \cap A)}{P(A)}, \text{ se } P(A) \neq 0 \quad (2.5)$$

Tratando-se de dois eventos permutáveis (seção 2.3.1.3), então:

$$P(A|B)P(B) = P(A \cap B) = P(B \cap A) = P(B|A)P(A) \quad (2.6)$$

Como:

$$P(A \cap B) = P(B|A)P(A) \quad (2.7)$$

Então, o Teorema de Bayes pode ser assim representado:

$$P(A|B) = \frac{P(B|A)}{P(B)}P(A) \quad (2.8)$$

A equação (2.8) é também chamada de **forma simples** do Teorema de Bayes

onde:

- $P(A)$ é a probabilidade *a priori*;
- $P(A|B)$ é a probabilidade *a posteriori*;
- $P(B|A)$ é a verossimilhança de B, dado A;
- $\frac{P(B|A)}{P(B)}$ é o fator de atualização das informações sobre o evento A.

Quando não se conhece o valor de $P(B)$, pode-se reescrever o valor dessa probabilidade em função de outras probabilidades de caráter mutuamente exclusivo (seção 2.3.1.4). Supondo-se que todo o espaço Ω possa ser dividido entre duas afirmações ou hipóteses antagônicas (e.g., no contexto do presente trabalho, hipótese da promotoria e a hipótese da defesa), o Teorema de Bayes pode então ser expresso como:

$$P(A|B) = \frac{P(B|A)P(A)}{P(B|A)P(A) + P(B|A^c)P(A^c)}P(A) \quad (2.9)$$

A equação (2.9) é também chamada de **forma alternativa** do Teorema de Bayes onde:

- $P(A)$ é a probabilidade *a priori*;
- $P(A^c) = 1 - P(A)$ é a probabilidade de que o conhecimento *a priori* seja falso;
- $P(B|A)$ é o grau de crença em B, dado que a proposição A é verdadeira. Também chamado de probabilidade condicional ou verossimilhança;
- $P(B|A^c)$ é o grau de crença em B, dado que a proposição A é falsa;
- $P(A|B)$ é a probabilidade *a posteriori*.

2.3.3. Desvendando o Dilema de Monty Hall usando o Teorema de Bayes

Uma vez definidas as equações do Teorema de Bayes (eqs. (2.8) e (2.9)), é possível retomar o Dilema de Monty Hall, apresentado na seção 2.2.1, e descobrir a melhor estratégia para resolvê-lo.

Inicialmente, foram apresentadas três portas: **uma com prêmio e duas sem prêmio**. Definindo-se:

- Evento A: o participante escolher a porta com o prêmio
- Evento B: o apresentador escolher a porta sem o prêmio

A probabilidade do participante acertar a porta com o prêmio sem nenhuma informação adicional é de uma chance em três ($P(A) = \frac{1}{3}$). Aqui não há nada de novo.

Agora, ao abordar as probabilidades que envolvem o apresentador, fica evidente **a principal vantagem do uso do Teorema de Bayes**. O apresentador tem uma informação privilegiada: ele sabe onde se encontra o prêmio, assim:

- a probabilidade do apresentador escolher uma porta sem prêmio antes do participante escolher uma porta é de 100 % ($P(B) = \frac{100}{100} = 1$).
- a probabilidade do apresentador escolher uma porta sem prêmio depois que o participante escolheu a porta com o prêmio é também de 100 % ($P(B|A) = \frac{100}{100} = 1$).

Substituindo-se os valores na equação (2.8) (**forma simples do Teorema de Bayes**), tem-se:

$$P(A|B) = \frac{P(B|A)}{P(B)}P(A) = \frac{1}{1} \cdot \frac{1}{3} = \frac{1}{3} \quad (2.10)$$

Interpretação do resultado acima: a probabilidade *a posteriori* (evento A dado B) de acertar a porta com o prêmio mantendo a escolha inicial de porta (evento A) dado que o apresentador abriu uma porta sem prêmio (evento B) continua sendo de 1

em 3, ou seja, a chance do prêmio estar na porta que sobrou é igual à probabilidade complementar de $\frac{2}{3}$. Assim, a melhor estratégia consiste em aproveitar a oportunidade dada pelo apresentador e sempre mudar de porta.

O fato que facilmente passa despercebido à intuição humana sobre o problema, ou seja, de que o apresentador tem uma informação privilegiada já que ele NUNCA vai escolher a porta com o prêmio, não passa despercebido ao cálculo por meio do Teorema de Bayes.

É interessante notar o que aconteceria com esse resultado no caso hipotético em que o apresentador não tivesse qualquer informação privilegiada sobre a posição do prêmio. Os eventos e as probabilidades seriam assim definidos:

- Evento A: o participante escolher a porta com o prêmio
- Evento B: o apresentador escolher a porta sem o prêmio

A probabilidade do participante acertar a porta com o prêmio sem nenhuma informação adicional continuaria sendo de uma chance em três ($P(A) = \frac{1}{3}$).

A probabilidade do apresentador escolher uma porta sem prêmio antes do participante escolher uma porta, já que ele não tem qualquer informação privilegiada, seria $P(B) = \frac{2}{3}$.

A probabilidade do apresentador escolher uma porta sem prêmio depois que o participante escolheu a porta com o prêmio (só restariam as duas portas sem prêmio) seria de 100 % ($P(B|A) = \frac{100}{100} = 1$).

Substituindo-se os valores na equação (2.8), tem-se:

$$P(A|B) = \frac{P(B|A)}{P(B)}P(A) = \frac{1}{\frac{2}{3}} \cdot \frac{1}{3} = \frac{3}{2} \cdot \frac{1}{3} = \frac{1}{2} = 50 \% \quad (2.11)$$

Percebe-se que a intuição humana inicial mais comum sobre o problema – de que a probabilidade de vitória permanecendo com a porta inicialmente escolhida é

igual à probabilidade de vitória caso opte-se pela troca para a porta restante – carece justamente da inclusão do fato de que o apresentador detém informação privilegiada sobre a posição em que se encontra o prêmio. A correta sistematização por meio do Teorema de Bayes não sofre dessa deficiência de cálculo.

2.3.4. Aplicação do Teorema de Bayes no contexto epidemiológico

O poder do formalismo do Teorema de Bayes pode ser também observado na interpretação de resultados de exames laboratoriais em um contexto epidemiológico.

Para interpretar o resultado de um teste para alguma doença, em geral, são necessárias três informações *a priori* (KAWAMURA, 2002):

1. Sensibilidade percentual do exame: é a probabilidade de um indivíduo avaliado e doente ter seu teste alterado (positivo);
2. Especificidade percentual do exame: é a probabilidade de um indivíduo avaliado e normal ter seu teste normal (negativo);
3. Prevalência populacional da doença: é a fração de indivíduos doentes na população total avaliada.

Para fins do presente exemplo, assumir-se-á o caso de um exame hipotético para COVID-19 com altíssimas taxas de sensibilidade (99 %) e de especificidade (99 %). Resta entender o papel da prevalência da doença e sua relação com a delimitação do subgrupo populacional em função de sintomas (anamnese) ou exposição a contágio.

A prevalência populacional de uma doença é um dado *a priori* que é construído ao longo da evolução de uma epidemia. Com os dados de 25 de maio de 2020 (WORLDMETERS.INFO, 2020), tinha-se que a população do Brasil era de 212 405 664 habitantes, havia 370 060 casos de COVID-19 confirmados e, até aquele instante, haviam sido realizados 735 224 testes. Acrescente-se ainda, a informação *a priori* de que, no Brasil, os testes estavam sendo majoritariamente realizados naqueles:

- que tivessem sintomas característicos de COVID-19;
- que tivessem tido contato com indivíduos contaminados por COVID-19;
- que fizessem parte de categorias profissionais com maior exposição ao contágio.
(ex.: profissionais de saúde e de segurança)

Com base nessas informações, há duas prevalências importantes a serem compreendidas:

- CASO 1: na população que apresenta algum sintoma ou tem histórico de exposição ao vírus Sars-CoV-2, a prevalência até aquele momento era de:
 $\frac{370060}{735224} \approx 50,33 \% \approx 50 \%$;
- CASO 2: na população geral, a prevalência da COVID-19, até aquele momento era de: $\frac{370060}{212405664} \approx 0,17 \% \approx 0,2 \%$

2.3.4.1. CASO 1: interpretação de exame em indivíduo de subgrupo populacional COM sintomas ou COM histórico de exposição ao vírus Sars-CoV-2

Suponha que 100000 indivíduos do CASO 1 (população que apresenta algum sintoma ou tem histórico de exposição ao vírus Sars-CoV-2) façam o teste para COVID-19 do exemplo hipotético em análise. Se um indivíduo tem seu resultado alterado (positivo), qual a probabilidade de que esse resultado signifique realmente que ele está doente? E se um indivíduo tem seu resultado inalterado (negativo), qual a probabilidade de que esse indivíduo esteja realmente saudável?

Dada a prevalência de aproximadamente 50 % nesse subgrupo populacional, imagina-se que 50000 têm a doença (+) e 50000 não têm a doença (-).

A sensibilidade do exame (99 %) permite inferir que haverá dentre os 50000 doentes (+): 49500 com resultado positivo (+ +) e 500 com resultado negativo (+ -).

A especificidade do exame (99 %) permite inferir que haverá dentre os 50000

indivíduos saudáveis (-): 49 500 com resultado negativo (- -) e 500 com resultado positivo (- +).

Ou seja, estima-se um total de 50 000 resultados positivos que serão compostos por: 49 500 doentes com resultado positivo (+ +) e 500 indivíduos saudáveis com resultado positivo (- +). Além disso, estima-se um total de 50 000 resultados negativos que serão compostos por: 500 doentes com resultado negativo (+ -) e 49 500 indivíduos saudáveis com resultado negativo (- -). Desses valores, é possível obter as taxas de falsos positivos (F.P.) e falsos negativos (F.N.):

$$\text{F.P.} = \frac{500}{50000} = 1 \% \quad (2.12)$$

$$\text{F.N.} = \frac{500}{50000} = 1 \% \quad (2.13)$$

O problema ora solucionado através de um exemplo numérico, pode ser diretamente resolvido aplicando-se o Teorema de Bayes.

- **Evento A: o indivíduo testado tem realmente COVID-19**
- **Evento B: o resultado de seu exame foi positivo**
- Probabilidade de um indivíduo ter COVID-19 *a priori* na porção da população que apresentou sintomas: $P(A) = 0,50$
- Probabilidade de um resultado dar positivo dado que o indivíduo tem COVID-19 (sensibilidade do teste): $P(B|A) = 0,99$
- Probabilidade de um resultado dar positivo *a priori*: $P(B) = \text{desconhecido}$

Como $P(B)$ é desconhecido, não é possível empregar a **forma simples** do Teorema de Bayes (eq. (2.8)). Entretanto, observe-se que o problema têm, intrinsecamente, um par de probabilidades de caráter mutuamente exclusivo:

- $P(A) = 0,50$ é a probabilidade de um indivíduo ter COVID-19 *a priori* na porção da

população que apresentou sintomas

- $P(A^c) = 1 - P(A) = 0,50$ é a probabilidade de um indivíduo não ter COVID-19 *a priori* na porção da população que apresentou sintomas

e sabendo-se ainda que:

- $P(B|A^c) = 0,01$ é a probabilidade de um resultado dar positivo dado que o indivíduo não tem COVID-19 (probabilidade complementar à especificidade)

então, é possível reescrever $P(B)$ como:

$$P(B) = P(B|A)P(A) + P(B|A^c)P(A^c) \quad (2.14)$$

Viabiliza-se, assim, o uso da **forma alternativa** do Teorema de Bayes (eq. (2.9)) para obter a probabilidade de que o indivíduo esteja realmente com COVID-19 dado que seu exame foi positivo ($P(A|B)$):

$$\begin{aligned} P(A|B) &= \frac{P(B|A)}{P(B|A)P(A) + P(B|A^c)P(A^c)}P(A) = \\ &= \frac{0,99}{(0,99 \cdot 0,50) + (0,01 \cdot 0,50)}0,50 = \\ &= 0,99 = 99 \% \quad (\text{F.P.} = 1 \%) \end{aligned} \quad (2.15)$$

Nesse exemplo, há simetria numérica para a obtenção do índice de falsos negativos, como pode ser facilmente verificado pelo leitor, definindo-se o evento A como “o indivíduo testado realmente não tem COVID-19” e o evento B como “o resultado de seu exame foi negativo”.

Nota-se, nesse exemplo, que estava correta a intuição inicial de que as probabilidades correspondiam aos valores de sensibilidade e especificidade. A conclusão é de que, quando a prevalência de uma doença em um subgrupo populacional é da ordem de 50 %, a confiabilidade do resultado é idêntica aos valores de sensibilidade e especificidade do exame. Fazendo-se a analogia ao Dilema de Monty Hall, a preva-

lência faz o papel do apresentador que, neste exemplo, não tem qualquer informação privilegiada antes da realização do exame (seção 2.3.3). Nesse caso, apenas por coincidência, é correta a concepção que temos sobre a qualidade dos exames ser diretamente proporcional a seus valores de especificidade e sensibilidade.

2.3.4.2. CASO 2: interpretação de exame em indivíduo de subgrupo populacional SEM sintomas ou SEM histórico de exposição ao vírus Sars-CoV-2

Suponha agora que 100000 indivíduos do CASO 2 (população geral que não apresenta qualquer sintoma ou que não tenha qualquer histórico de exposição ao vírus Sars-CoV-2) façam o teste para COVID-19 do exemplo hipotético em análise. Se um indivíduo tem seu resultado alterado (positivo), qual a probabilidade de que esse resultado signifique realmente que ele está doente? E se um indivíduo tem seu resultado inalterado (negativo), qual a probabilidade de que esse indivíduo esteja realmente saudável?

A prevalência agora é de aproximadamente 0,2 % nesse subgrupo populacional o que permite inferir que 200 têm a doença (+) e 99800 não têm a doença (-).

A sensibilidade do exame (99 %) permite inferir que haverá dentre os 200 doentes (+): 198 com resultado positivo (+ +) e 2 com resultado negativo (+ -).

A especificidade do exame (99 %) permite inferir que haverá dentre os 99800 indivíduos saudáveis (-): 98802 com resultado negativo (- -) e 998 com resultado positivo (- +).

Ou seja, estima-se um total de 1196 resultados positivos que serão compostos por: 198 doentes com resultado positivo (+ +) e 998 indivíduos saudáveis com resultado positivo (- +). Além disso, estima-se um total de 98804 resultados negativos que serão compostos por: 2 doentes com resultado negativo (+ -) e 98802 indivíduos saudáveis com resultado negativo (- -). Desses valores, é possível obter as taxas de falsos

positivos (F.P.) e falsos negativos (F.N.):

$$\text{F.P.} = \frac{998}{1196} \approx 83,44 \% \quad (2.16)$$

$$\text{F.N.} = \frac{2}{98804} \approx 0,002 \% \quad (2.17)$$

Já é possível evidenciar que a intuição inicial de que as probabilidades correspondam aos valores de sensibilidade e especificidade é incorreta. Interpretando-se o resultado acima, naquele instante da pandemia de COVID-19, a aplicação de um exame com altíssimas taxas de sensibilidade e especificidade (hipotéticas, posto que nenhum exame atingia esses limiares até a data dos dados aqui empregados (WORLDOMETERS.INFO, 2020)) em uma população sem risco prévio de contágio, faria com que a probabilidade de um resultado positivo ser falso fosse de 83,44 %, enquanto os resultados negativos teriam maior força dado que a probabilidade de falsos negativos seria de apenas 0,002 %.

A solução direta desse problema passa pela aplicação do Teorema de Bayes.

2.3.4.2.1 Obtenção do índice de falsos negativos

Sejam os termos:

- **Evento A: o indivíduo testado realmente não tem COVID-19**
- **Evento B: o resultado de seu exame foi negativo**
- Probabilidade de um indivíduo não ter COVID-19 *a priori* na porção da população que não apresentou sintomas, nem teve exposição ao contágio: $P(A) = 0,998$
- Probabilidade de um indivíduo ter COVID-19 *a priori* na porção da população que não apresentou sintomas nem teve exposição ao contágio: $P(A^c) = 1 - P(A) = 0,002$

- Probabilidade de um resultado dar negativo dado que o indivíduo não tem COVID-19 (especificidade do teste): $P(B|A) = 0,99$
- Probabilidade de um resultado dar negativo dado que o indivíduo tem COVID-19 (probabilidade complementar à sensibilidade): $P(B|A^c) = 0,01$

Substituindo-se esses valores na **forma alternativa** do Teorema de Bayes (eq. (2.9)):

$$\begin{aligned}
 P(A|B) &= \frac{P(B|A)}{P(B|A)P(A) + P(B|A^c)P(A^c)}P(A) = \\
 &= \frac{0,99}{(0,99 \cdot 0,998) + (0,01 \cdot 0,002)}0,998 = \\
 &\approx 0,99998 = 99,998 \% \quad (\text{F.N.} = 0,002 \%)
 \end{aligned} \tag{2.18}$$

Portanto, quando um indivíduo faz parte de um subgrupo populacional em que a prevalência de COVID-19 é baixa (no exemplo é da ordem de 0,2 %), a probabilidade de que essa pessoa realmente não tenha COVID-19, dado que seu exame deu resultado negativo, é cerca de 99,998 %. A interpretação desse resultado não deve gerar dúvidas quanto a sua confiabilidade ao médico que solicitou o exame.

2.3.4.2.2 Obtenção do índice de falsos positivos

Sejam os termos:

- **Evento A: o indivíduo testado tem realmente COVID-19**
- **Evento B: o resultado de seu exame foi positivo**
- Probabilidade de um indivíduo ter COVID-19 *a priori* na porção da população que não apresentou sintomas, nem teve exposição ao contágio: $P(A) = 0,002$
- Probabilidade de um indivíduo não ter COVID-19 *a priori* na porção da população que não apresentou sintomas nem teve exposição ao contágio: $P(A^c) = 1 - P(A) = 0,998$

- Probabilidade de um resultado dar positivo dado que o indivíduo tem COVID-19 (sensibilidade do teste): $P(B|A) = 0,99$
- Probabilidade de um resultado dar positivo dado que o indivíduo não tem COVID-19 (probabilidade complementar à especificidade): $P(B|A^c) = 0,01$

Substituindo-se esses valores na **forma alternativa** do Teorema de Bayes (eq. (2.9)):

$$\begin{aligned}
 P(A|B) &= \frac{P(B|A)}{P(B|A)P(A) + P(B|A^c)P(A^c)}P(A) = \\
 &= \frac{0,99}{(0,99 \cdot 0,002) + (0,01 \cdot 0,998)}0,002 \approx \\
 &\approx 0,1656 = 16,56 \% \quad (\text{F.P.} = 83,44 \%)
 \end{aligned}
 \tag{2.19}$$

Portanto, quando um indivíduo faz parte de um subgrupo populacional em que a prevalência de COVID-19 é baixa (no exemplo é da ordem de 0,2 %), a probabilidade de que essa pessoa tenha realmente COVID-19, dado que seu exame deu resultado positivo, é de apenas 16,56 %. Nesse caso, a interpretação desse resultado deve gerar dúvidas quanto a sua confiabilidade ao médico que solicitou o exame. Um novo exame deverá ser solicitado e sua interpretação deverá passar por uma aplicação recursiva do Teorema de Bayes (seção 2.3.4.2.3).

Neste ponto, cabe observar que, quando a prevalência de uma doença em um subgrupo populacional é conhecida (neste exemplo, 0,002, a confiabilidade do resultado deve ser calculada à luz do Teorema de Bayes. Fazendo-se a analogia ao Dilema de Monty Hall (2.3.3), a prevalência faz agora o papel de um apresentador que tem informação privilegiada antes da realização do exame. Nesses casos, torna-se equivocada a intuição inicial de que a probabilidade de um exame positivo ou negativo corresponder à verdade é proporcional à sensibilidade e à especificidade do próprio exame, respectivamente.

É importante ressaltar também que a prevalência populacional da doença não é

um dado fixo. A análise de qualquer resultado de exame em contexto epidemiológico depende da constante readequação da prevalência populacional à realidade temporal da evolução da epidemia e à definição de qual é o subgrupo populacional no qual o indivíduo testado está inserido.

2.3.4.2.3 Aplicação recursiva do Teorema de Bayes

Suponha que o mesmo indivíduo do exemplo da seção 2.3.4.2.2 seja novamente testado a fim de dirimir a dúvida sobre a interpretação de seu resultado positivo. Nessa condição – e frise-se que esta etapa é ponto de costumeira confusão dentre aqueles que usam o Teorema de Bayes –, a interpretação desse novo resultado precisará da **aplicação recursiva do Teorema de Bayes**.

O valor calculado da probabilidade de que essa pessoa tenha realmente COVID-19, dado que seu exame deu resultado positivo, é de apenas 16,56 %. No cálculo feito à equação (2.19), esse valor correspondia ao termo $P(A|B)$, ou probabilidade *a posteriori* de que o indivíduo testado tem realmente COVID-19 (evento A) dado que o resultado de seu exame foi positivo (evento B).

Pois bem, na reaplicação do exame, o que antes era probabilidade *a posteriori* passa a ser a informação prévia que se tem sobre a condição do indivíduo. Ou seja, na reaplicação do exame o valor anteriormente obtido de $P(A|B)$ passa a ser a informação *a priori* $P(A)$.

Supondo que o resultado de seu novo exame tenha sido positivo, então, a reaplicação do Teorema de Bayes consistirá em recalcular a equação (2.19), substituindo-se o novo valor *a priori* $P(A)$. Dessa forma:

$$\begin{aligned}
 P(A|B) &= \frac{P(B|A)}{P(B|A)P(A) + P(B|A^c)P(A^c)}P(A) = \\
 &= \frac{0,99}{(0,99 \cdot \mathbf{0,1656}) + (0,01 \cdot \mathbf{0,8344})} \mathbf{0,1656} \approx \\
 &\approx 0,9516 = 95,16 \% \quad (\text{F.P.} = 4,84 \%)
 \end{aligned} \tag{2.20}$$

Ou seja, nas condições do exemplo proposto, mesmo em um indivíduo que não fazia parte do grupo de risco (CASO 2), dois resultados positivos serão suficientes para obter uma probabilidade de 95,16 % de que o indivíduo realmente tenha a doença.

Agora, caso o resultado do segundo exame retorne negativo, será necessário aplicar a recursividade do Teorema de Bayes na equação (2.18) da seção 2.3.4.2.1, alterando-se o valor da probabilidade *a priori* para o evento *A*, ali definido como “o indivíduo testado realmente não tem COVID-19”, como sendo o valor de falso positivo obtido a partir da interpretação do resultado do primeiro exame. Ou seja, agora, a probabilidade *a priori* de que o indivíduo testado não tenha COVID-19, que antes era $P(A) = 99,8 \%$, passa a ser $P(A) = 83,44 \%$. O novo cálculo segue abaixo:

$$\begin{aligned}
 P(A|B) &= \frac{P(B|A)}{P(B|A)P(A) + P(B|A^c)P(A^c)}P(A) = \\
 &= \frac{0,99}{(0,99 \cdot \mathbf{0,8344}) + (0,01 \cdot \mathbf{0,1656})} \mathbf{0,8344} = \\
 &\approx 0,9980 = 99,80 \% \quad (\text{F.N.} = 0,20 \%)
 \end{aligned} \tag{2.21}$$

Portanto, nas condições do exemplo proposto, para um indivíduo que não faz parte do grupo de risco (CASO 2), um resultado positivo seguido de um resultado negativo permitiriam inferir a probabilidade de 99,80 % de que o indivíduo realmente não tenha a doença.

2.3.5. Razão de Verossimilhanças

No contexto jurídico, uma evidência (E) é interpretada atendendo a duas visões tipicamente antagônicas:

- a Hipótese da Promotoria (H_P)
- e a Hipótese da Defesa (H_D)

Em termos do Teorema de Bayes, tem-se que a probabilidade *a posteriori* da hipótese da promotoria estar certa dado que a evidência é verdadeira ($P(H_P|E)$) é dada por:

$$P(H_P|E) = \frac{P(E|H_P)}{P(E)}P(H_P) \quad (2.22)$$

onde:

- $P(E|H_P)$ é a probabilidade da evidência ser verdadeira dado que a hipótese da promotoria está certa
- $P(E)$ é a probabilidade a priori da evidência ser verdadeira
- $P(H_P)$ é a probabilidade a priori da hipótese da promotoria estar certa.

Analogamente ao acima explicado, é possível definir a probabilidade *a posteriori* da hipótese da defesa estar certa dado que a evidência é verdadeira ($P(H_D|E)$) como:

$$P(H_D|E) = \frac{P(E|H_D)}{P(E)}P(H_D) \quad (2.23)$$

Porém, nas formulações acima, há elementos componentes da análise bayesiana proposta que são intrinsecamente intangíveis. É impossível saber $P(E)$, pois necessariamente, seria preciso testar a evidência por métodos comparativos sujeitos a incertezas que afetariam essa avaliação *a priori* sobre a evidência. Além disso, $P(H_P)$ – ou $P(H_D)$ – corresponde à confiança que o Promotor de Justiça – ou o Defensor – tem

sobre a força de sua hipótese e isso é fruto de considerações pessoais das partes.

Ou seja, estabelecer a força da hipótese diante da evidência – $P(H_P|E)$ – está além das incumbências do perito criminal. A obtenção dos resultados $P(H_P|E)$ e $P(H_D|E)$ requer pesar todas as evidências apresentadas e comparar os valores relativos entre elas. Este é o papel de quem julga.

Felizmente, existe uma solução alternativa para a apresentação dos resultados de perícias técnico-científicas. Dividindo-se a equação (2.22) pela equação (2.23) tem-se:

$$\frac{P(H_P|E)}{P(H_D|E)} = \frac{P(E|H_P)}{P(E|H_D)} \cdot \frac{P(H_P)}{P(H_D)} \quad (2.24)$$

As partes componentes da equação (2.24) podem ser assim nomeadas:

- $\frac{P(H_P|E)}{P(H_D|E)}$ é a razão de probabilidades *a posteriori*
- $\frac{P(H_P)}{P(H_D)}$ é a razão de probabilidades *a priori*
- $\frac{P(E|H_P)}{P(E|H_D)}$ é a razão de verossimilhanças.

A razão de probabilidades *a priori* representaria a crença do julgador no peso relativo das duas hipóteses levantadas pela Promotoria e pela Defesa antes de que qualquer evidência tenha sido apresentada.

A razão de probabilidades *a posteriori* consiste no valor ajustado da razão de probabilidades *a priori* diante das evidências. Nos moldes aqui delineados, este cálculo seria privativo ao julgador posto que isso requereria incluir informações e suposições de fontes além da avaliação científica da evidência.

A razão de verossimilhanças (LR ²⁷) mostra-se, assim, como a forma ideal para que o Perito Criminal apresente seus resultados em testes que se baseiem em análises de evidências em hipóteses antagônicas (MORRISON, 2009):

²⁷Do inglês, *Likelihood Ratio*.

$$LR = \frac{P(E|H_P)}{P(E|H_D)} \quad (2.25)$$

Em algumas publicações, a notação para a razão de verossimilhanças é mais completa de modo a evidenciar os fatos de que a medição se refere ao peso da evidência e de que a relação entre numerador e divisor empregada é entre H_P e H_D . Essa notação mais completa é a que segue:

$$LR_{H_P, H_D}(E) = \frac{P(E|H_P)}{P(E|H_D)} \quad (2.26)$$

No presente texto, optou-se pela forma da equação (2.25).

2.3.5.1. Aplicação recursiva da Razão de Verossimilhanças

Uma das vantagens da apresentação de resultados por razão de verossimilhanças está na simplicidade de sua aplicação recursiva, quando da necessidade de se avaliar o impacto de diversos resultados periciais distintos.

Retomando-se as equações (2.24) e (2.25), o impacto de um primeiro valor de razão de verossimilhanças sobre os valores *a posteriori* pode ser descrito como:

$$\frac{P_1(H_P|E)}{P_1(H_D|E)} = LR_1 \cdot \frac{P(H_P)}{P(H_D)} \quad (2.27)$$

O impacto de uma novo resultado de razão de verossimilhanças pode ser calculado por meio da aplicação recursiva em que, agora, o valor *a priori* corresponde ao resultado *a posteriori* da equação (2.27). Ou seja:

$$\frac{P_2(H_P|E)}{P_2(H_D|E)} = LR_2 \cdot \left(LR_1 \cdot \frac{P(H_P)}{P(H_D)} \right) \quad (2.28)$$

Uma terceira interação com base em um outro resultado de LR, seria:

$$\frac{P_3(H_P|E)}{P_3(H_D|E)} = LR_3 \cdot \left[LR_2 \cdot \left(LR_1 \cdot \frac{P(H_P)}{P(H_D)} \right) \right] \quad (2.29)$$

Ou seja, de modo geral, a interpretação de $i = 1 \dots n$ resultados apresentados em termos de n razões de verossimilhanças consiste simplesmente no produtório desses valores:

$$\frac{P(H_P|E)}{P(H_D|E)} = \left(\prod_{i=1}^n LR_i \right) \cdot \frac{P(H_P)}{P(H_D)} \quad (2.30)$$

2.3.5.2. Abordagem gráfica para obtenção da Razão de Verossimilhanças

Outra forma de ser calculada a Razão de Verossimilhanças (LR) em um contexto de comparação forense – especialmente características biométricas – é a abordagem gráfica (ALI; SPREEUWERS; VELDHUIS, 2012, p. 7-9).

Retomando-se a equação (2.25), é possível adaptá-la para incluir uma população relevante – definida como uma população de indivíduos com características semelhantes àquelas do indivíduo questionado – também conhecida como informação de *background* no modelo UBM (ver sec. 3.1.3).

$$LR = \frac{P(E|H_P, I)}{P(E|H_D, I)} \quad (2.31)$$

onde H_P e H_D são as hipóteses da promotoria e da defesa, como anteriormente definidos, E é a informação forense – ou evidência – e I é a informação de *background* relativa ao caso analisado.

O numerador da equação (2.31) é obtido calculando-se a intravariabilidade (WSV²⁸) do material analisado, enquanto o denominador é calculado por meio da intervariabilidade (BSV²⁹) entre o material analisado e a população relevante. A obtenção

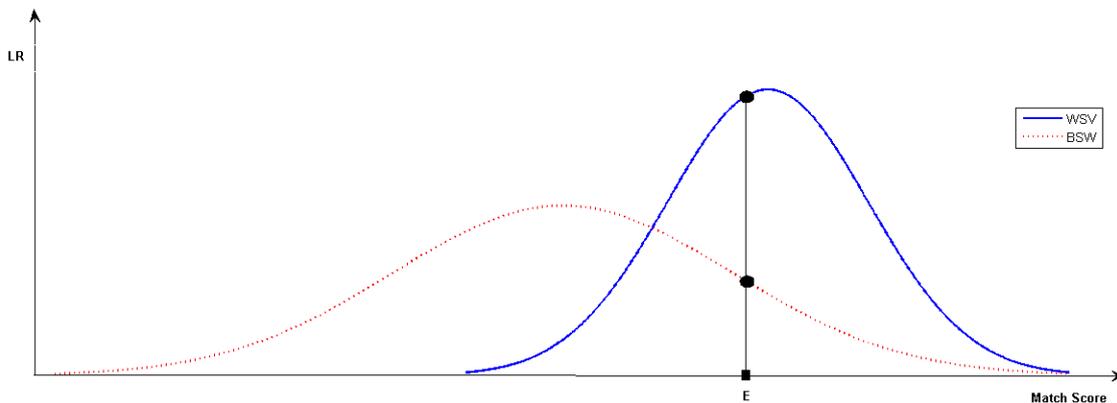
²⁸Do inglês, *within-source variability*.

²⁹Do inglês, *between-source variability*.

do valor de LR depende ainda do escore E , que é calculado comparando-se o vestígio do caso analisado com a amostra de fonte conhecida.

Uma vez obtidos os citados escores – WSV , BSV e E –, calcula-se LR dividindo-se a altura de H_P pela altura de H_D no ponto E , como ilustrado na figura 2.7³⁰.

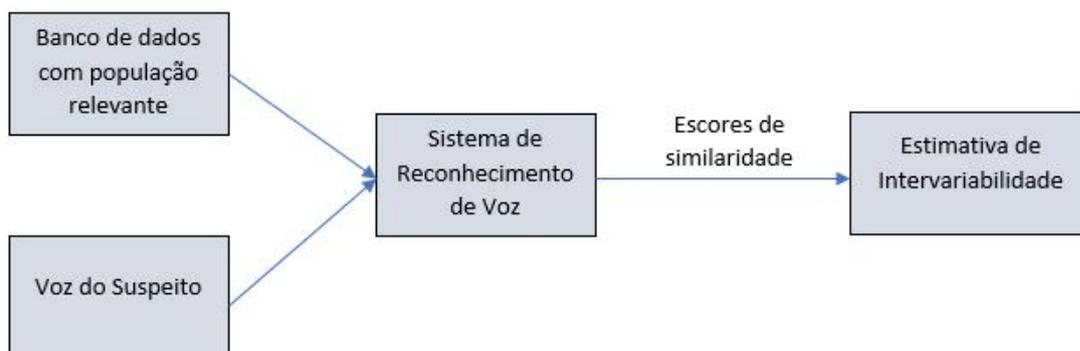
Figura 2.7. – Abordagem gráfica para cálculo da Razão de Verossimilhanças.



A curva sólida mostra a distribuição de escores H_P , enquanto a curva pontilhada mostra a distribuição de escores H_D . LR é calculada pela divisão da altura de H_P por H_D no ponto E . Fonte: (ALI; SPREEUWERS; VELDHUIS, 2012, p. 8, tradução minha).

A estrutura esquemática exemplificativa para a apresentação de resultados por Razão de Verossimilhanças em um sistema de comparação de vozes consta das figuras 2.8 a 2.10.

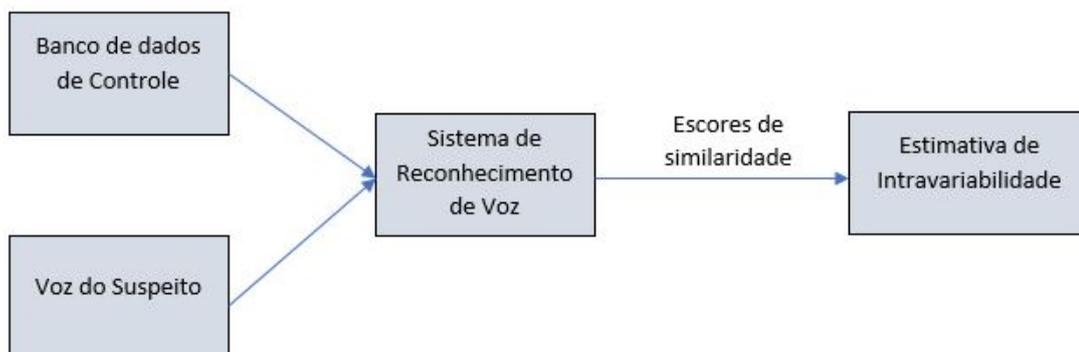
Figura 2.8. – Estimativa da Intervariabilidade (BSV).



Fonte: (ALI; SPREEUWERS; VELDHUIS, 2012, p. 9, tradução e adaptação minhas).

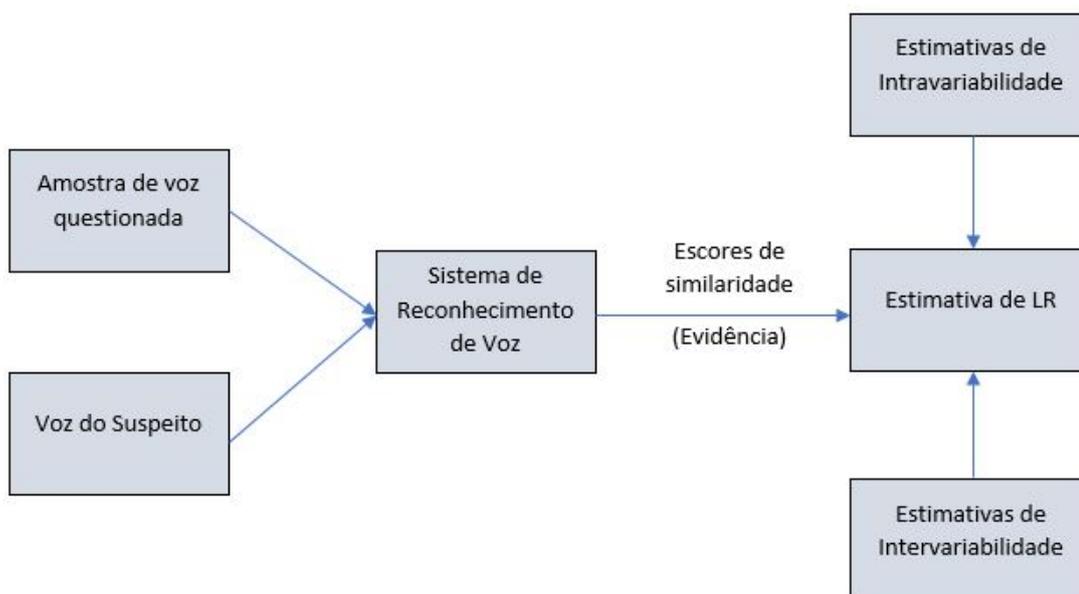
³⁰The solid curve shows the distribution of H_p scores while the dotted curve represents the distribution of H_d scores. LR is calculated by dividing the height of H_p by H_d at point E .

Figura 2.9. – Estimativa da Intravariabilidade (WSV).



Fonte: (ALI; SPREEUWERS; VELDHUIS, 2012, p. 9, tradução e adaptação minhas).

Figura 2.10. – Cálculo da Razão de Verossimilhanças (LR).



Fonte: (ALI; SPREEUWERS; VELDHUIS, 2012, p. 9, tradução e adaptação minhas).

2.3.5.3. Apresentação dos resultados: LR ou $\log LR$

Como já explicado, o valor numérico da Razão de Verossimilhanças (LR) representa o número de vezes que uma hipótese H_1 é mais condizente com as evidências analisadas em comparação ao quão condizente outra hipótese alternativa H_2 é diante da mesma evidência. Pois bem, não é incomum que resultados de LR sejam da ordem de 100, 1000, 10000 ou mesmo valores superiores. Por esse motivo, também não é incomum encontrar esses resultados apresentados na forma logarítmica de base 10, posto que é usual ao cientista recorrer a essa escala para mostrar resultados multiplicativos.

Para ilustrar, observe a tabela de correspondência abaixo:

Tabela 2.2. – LR x $\log LR$: Comparativo entre as apresentações dos resultados de LR no modo fracionário e no modo logarítmico.

LR	$\log LR$	Interpretação
$10000/1$	+4	H_1 é 10000 vezes mais “forte” que H_2
$1000/1$	+3	H_1 é 1000 vezes mais “forte” que H_2
$100/1$	+2	H_1 é 100 vezes mais “forte” que H_2
$10/1$	+1	H_1 é 10 vezes mais “forte” que H_2
$1/1$	0	H_1 e H_2 têm a mesma “força”
$1/10$	-1	H_2 é 10 vezes mais “forte” que H_1
$1/100$	-2	H_2 é 100 vezes mais “forte” que H_1
$1/1000$	-3	H_2 é 1000 vezes mais “forte” que H_1
$1/10000$	-4	H_2 é 10000 vezes mais “forte” que H_1

Fonte: produção própria.

O objetivo primordial da apresentação de resultados periciais reside na clareza e inteligibilidade: é preciso considerar que os resultados não serão apreciados unicamente por um público técnico, mas também por leigos. A escala logarítmica, embora extremamente clara a cientistas, traduz ao leigo uma falsa linearidade aparente para grandezas que se relacionam de modo multiplicativo. É muito intuitivo ao leigo

a interpretação de que a “distância” relativa entre os valores de $\log LR$, por exemplo, $+4$ e $+2$ é de “duas unidades” – o que quer que isso signifique do ponto de vista do leigo –, portanto, o $\log LR = +3$ estaria “no meio do caminho” entre esses dois valores. Ocorre que, de fato, $\log LR = +3$ é 10 vezes maior do que $\log LR = +2$ e o $\log LR = +4$ é 100 vezes maior do que $\log LR = +2$, o que evidencia o problema interpretativo. Além disso, como se vê na tabela 2.2, a escala logarítmica poderia ser equivocadamente associada à escala de nove pontos proposta pela Polícia Sueca (ver sec. 2.1.4), em que a distribuição de notas igualmente espaçadas variando de -4 a $+4$ traz consigo uma interpretação subjetiva. Assim, fica evidente a necessidade de se apresentar resultados de Razão de Verossimilhanças em sua forma fracionária, não dando margem para interpretações alheias à realidade dos cálculos.

Nesse ponto em especial, diante do quanto exposto, teço a recomendação crítica de que resultados periciais produzidos por meio de confronto de hipóteses concorrentes e resultantes em valores de Razão de Verossimilhanças devam necessariamente ser apresentados no modo fracionário e acompanhados de sua interpretação.

2.3.5.4. Incertezas

A atenta compreensão da estrutura matemática da Razão de Verossimilhanças, até aqui explicada, permite inferir que a composição do valor de LR não traga consigo informação intrínseca sobre incertezas de medida. É possível notar também que a interpretação direta de LR como sendo a probabilidade de que uma hipótese seja verdadeira é algo incorreto – de fato, o valor de LR representa a quantidade de vezes que uma hipótese é mais provável de ser verdadeira em comparação a uma hipótese alternativa e diante da evidência analisada.

Como muito bem argumentado por Slooten (2020), não existe incerteza associada à Razão de Verossimilhanças. As incertezas somente podem ser computadas com relação às decisões ou ações baseadas em limiares pré-definidos de LR à luz das relações entre as hipóteses concorrentes e das probabilidades *a priori*.

Tomando-se por base as hipóteses concorrentes H_1 e H_2 e definindo-se:

- A_1 : a ação ou decisão que se toma caso H_1 seja verdadeira;
- A_2 : a ação ou decisão que se toma caso H_2 seja verdadeira;
- $P(A_2|H_1) = \alpha$: a probabilidade de se tomar a indesejada ação A_2 caso H_1 seja verdadeira;
- $P(A_1|H_2) = \beta$: a probabilidade de se tomar a indesejada ação A_1 caso H_2 seja verdadeira.

Dessas definições decorre que:

- $P(A_1|H_1) = 1 - \alpha$: é a probabilidade de se tomar a ação A_1 , dado que a H_1 seja verdadeira;
- $P(A_2|H_2) = 1 - \beta$: é a probabilidade de se tomar a ação A_2 , dado que H_2 seja verdadeira.

A referência (SLOOTEN, 2020) apresenta metodologia para a obtenção da probabilidade *a posteriori* de que H_1 esteja certa dado que se tomou a decisão A_1 :

$$P(H_1|A_1) = \frac{(1 - \alpha)(\beta - p_{A_1})}{p_{A_1}(\alpha + \beta - 1)} \quad (2.32)$$

Onde p_{A_i} é a proporção de decisões em que a ação A_i foi escolhida.

2.4. Abordagem bayesiana para apresentação de resultados periciais

Do tanto quanto exposto, é possível depreender que o Teorema de Bayes é aplicável em algumas situações práticas em que humanos poderiam ser levados a tirar conclusões subótimas devido a vieses cognitivos. Nesta seção, serão apresentados exemplos práticos no contexto jurídico, de apresentação de resultados periciais

empregando-se argumentação construída com base na Razão de Verossimilhanças.

2.4.1. Exemplo no contexto de análise de voz

Os peritos forenses devem atestar similaridades e dissimilaridades entre amostras de voz, mas não podem fazer afirmações categóricas (julgamentos) sobre as vozes em si. A ferramenta estatística largamente empregada para computar a força das evidências é a **Razão de Verossimilhança**.

Supondo o exemplo em que sejam:

- X : amostra de voz gravada (evidência)
- Y : amostra de voz obtida de um suspeito
- H_P : hipótese que X e Y são falados pela mesma pessoa
- H_D : hipótese que X e Y são falados por pessoas diferentes
- E : evidência forense observada (e.g. *pitch* médio entre X e Y diferem em 10 Hz)

A razão de verossimilhança será definida por:

$$LR = \frac{P(E|H_P)}{P(E|H_D)} \quad (2.33)$$

Na metade superior da tabela 2.3, foi mantida fixa a probabilidade de 80 % de similaridade entre a evidência e a amostra de um suspeito, enquanto foi sendo variada a probabilidade de que a evidência seja compatível com qualquer pessoa ao acaso no público geral. Quando há alta similaridade entre a evidência e a amostra de um suspeito e baixa similaridade entre a evidência e a voz de qualquer outra pessoa ao acaso, por exemplo, $P(E|H_D) = 10 \%$, a razão de verossimilhança é alta – nesse caso, $LR = 8$. Quando a voz analisada tem alta probabilidade de ter sido originada por uma pessoa qualquer, por exemplo, $P(E|H_D) = 80 \%$, o valor de LR é baixo, neste caso, vale 1. Uma razão de verossimilhança igual a 1, equivale a dizer que a voz analisada tem

iguais chances de ser do suspeito ou de qualquer outra pessoa.

Já na metade inferior da tabela 2.3, inverte-se o proposto: mantém-se fixa a probabilidade – alta – de que a amostra analisada tenha similaridades bastante genéricas que indiquem que ela provém de qualquer pessoa ao acaso, enquanto varia-se a probabilidade de que ela seja compatível com a amostra obtida de determinado suspeito. Por exemplo, uma $P(E|H_P) = 10\%$ contra uma $P(E|H_D) = 80\%$ leva a uma $LR = 0,125$ que apóia a hipótese da defesa de que diferentes falantes podem estar envolvidos no caso investigado.

A interpretação de valores de LR decimais é melhor entendida em termos de seus inversos. Por exemplo, quando $LR = 0,125$, depreende-se que as chances em favor da defesa são de 8 para 1, posto que $\frac{1}{0,125} = \frac{8}{1}$.

Tabela 2.3. – Razão de Verossimilhança.

$\frac{P(E H_P)}{P(E H_D)}$	$\frac{80\%}{10\%}$	$\frac{80\%}{20\%}$	$\frac{80\%}{40\%}$	$\frac{80\%}{80\%}$
LR	8	4	2	1
$\frac{P(E H_P)}{P(E H_D)}$	$\frac{10\%}{80\%}$	$\frac{20\%}{80\%}$	$\frac{40\%}{80\%}$	$\frac{80\%}{80\%}$
LR	0,125	0,25	0,5	1

Funcionamento da razão de verossimilhança (LR): $P(E|H_P)$ é a probabilidade de que uma evidência obtida seja de determinado suspeito e $P(E|H_D)$ é a probabilidade de que a evidência obtida seja de qualquer outra pessoa analisada ao acaso. Na parte superior da tabela, $P(E|H_P)$ foi fixado em um valor alto, enquanto $P(E|H_D)$ foi gradualmente aumentado. Nesse caso, LR varia de 8 a 1 o que impacta em um drástico enfraquecimento da hipótese da acusação (de que a voz obtida como evidência pertença ao suspeito). Na parte inferior, ocorre o inverso e vê-se o enfraquecimento da hipótese da acusação (de que a voz obtida como evidência pertença não pertença ao investigado). Fonte: (ROSE, 2002, cap. 4).

2.4.2. Exemplo: comparação entre objetos apreendidos e filmagens de um crime

Suponha-se um caso de comparação das vestimentas encontradas com um suspeito de um crime com aquelas presentes em um vídeo da ação criminosa em que

não seja possível ver os rostos dos envolvidos³¹.

A autoridade policial encontrou com os investigados por roubo uma calça jeans, um par de tênis e uma mochila parcialmente destruída por ação de fogo, conforme imagens 2.11, 2.12 e 2.13.

Figura 2.11. – Calça encaminhada para perícia.



Fonte: produção própria.

A ação criminosa dos dois envolvidos, um que aborda o condutor do veículo que será roubado e outra que segue na motocicleta, foi filmada conforme imagem 2.14.

O trabalho pericial passou pela comparação entre os valores RGB³² das cores dos objetos presentes nas filmagens e esses mesmos valores obtidos em fotos produzidas durante o levantamento pericial. Além disso, foram elencadas as características identificadoras dos objetos examinados e daqueles presentes nas filmagens. Esse levantamento consta das imagens 2.15 a 2.18.

³¹ As imagens aqui apresentadas referem-se a um caso real. Todos os elementos que possam servir para identificar o caso e/ou envolvidos foram retirados ou desfocados.

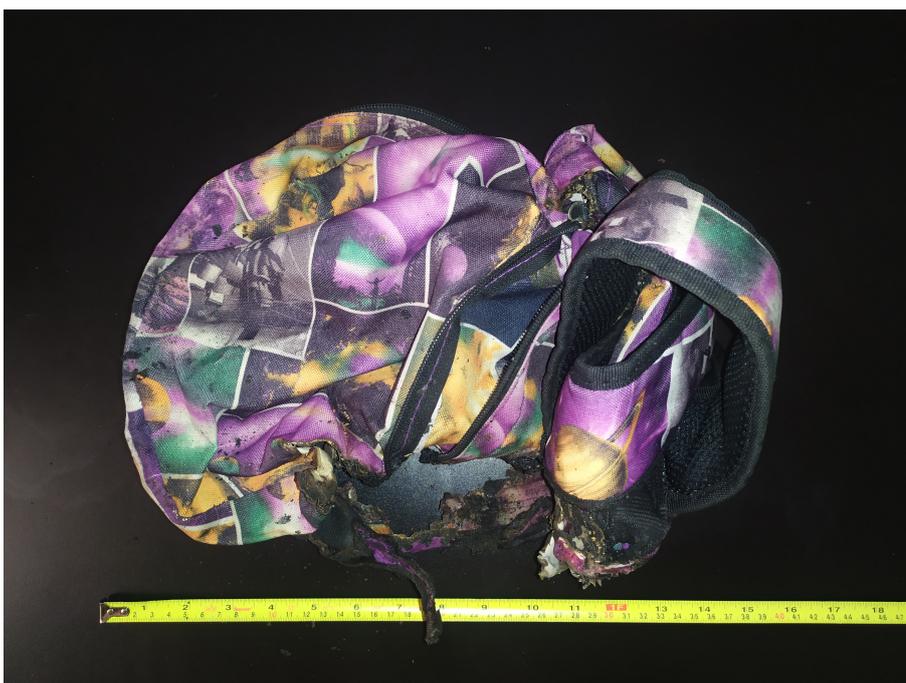
³² RGB é a abreviatura de um sistema de cores aditivas em que o Vermelho (*Red*), o Verde (*Green*) e o Azul (*Blue*) são combinados de várias formas de modo a reproduzir um largo espectro cromático (RGB, 2021).

Figura 2.12. – Par de tênis encaminhado para perícia.



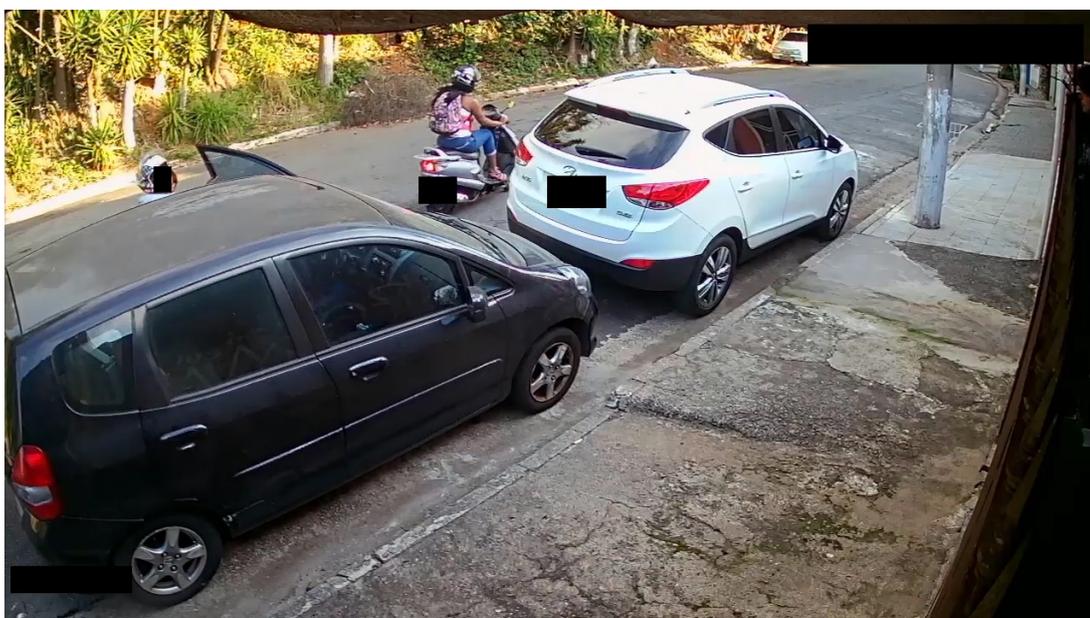
Fonte: produção própria.

Figura 2.13. – Mochila encaminhada para perícia.



Fonte: produção própria.

Figura 2.14. – Cena de filmagem de ação criminosa por câmera de segurança.



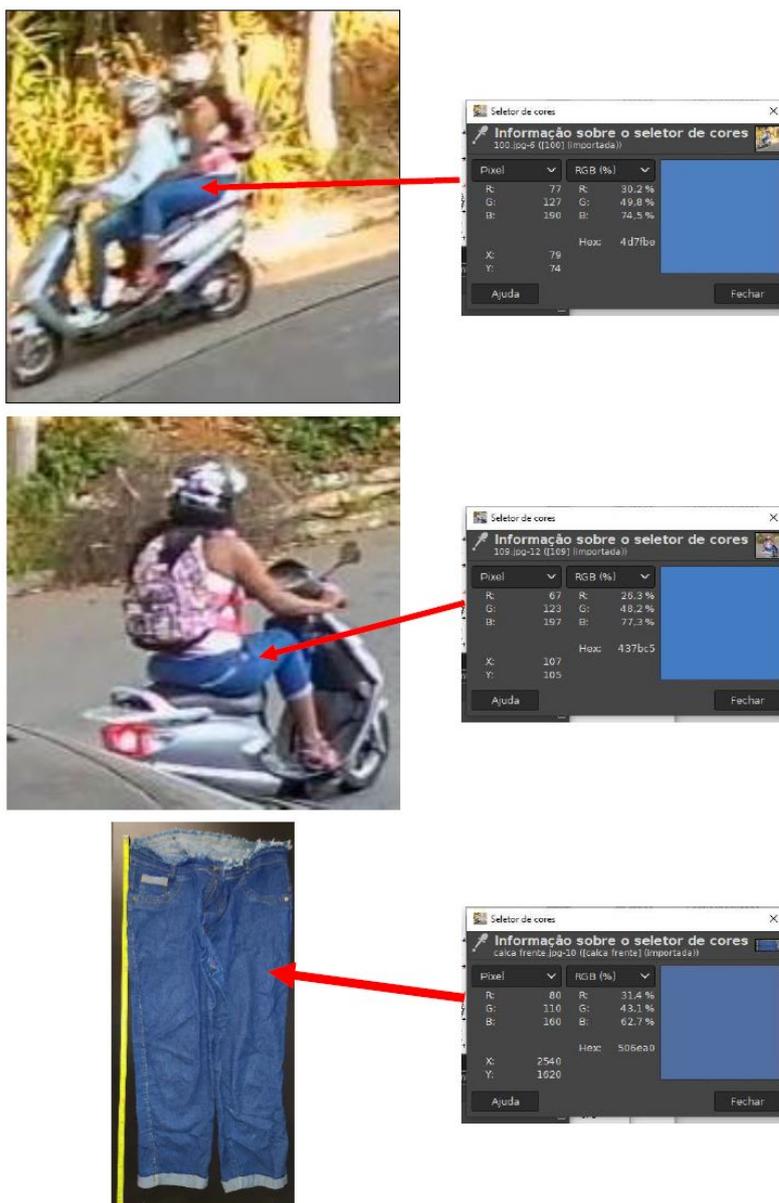
Fonte: produção própria adaptada de caso real. Descaracterizado para manutenção de sigilo.

Figura 2.15. – Detalhes da calça examinada.



A calça examinada apresenta 5 (cinco) detalhes coincidentes àqueles da calça utilizada por um dos indivíduos presentes da ação delituosa. Fonte: produção própria adaptada de caso real. Descaracterizado para manutenção de sigilo.

Figura 2.16. – Detalhes da calça examinada.



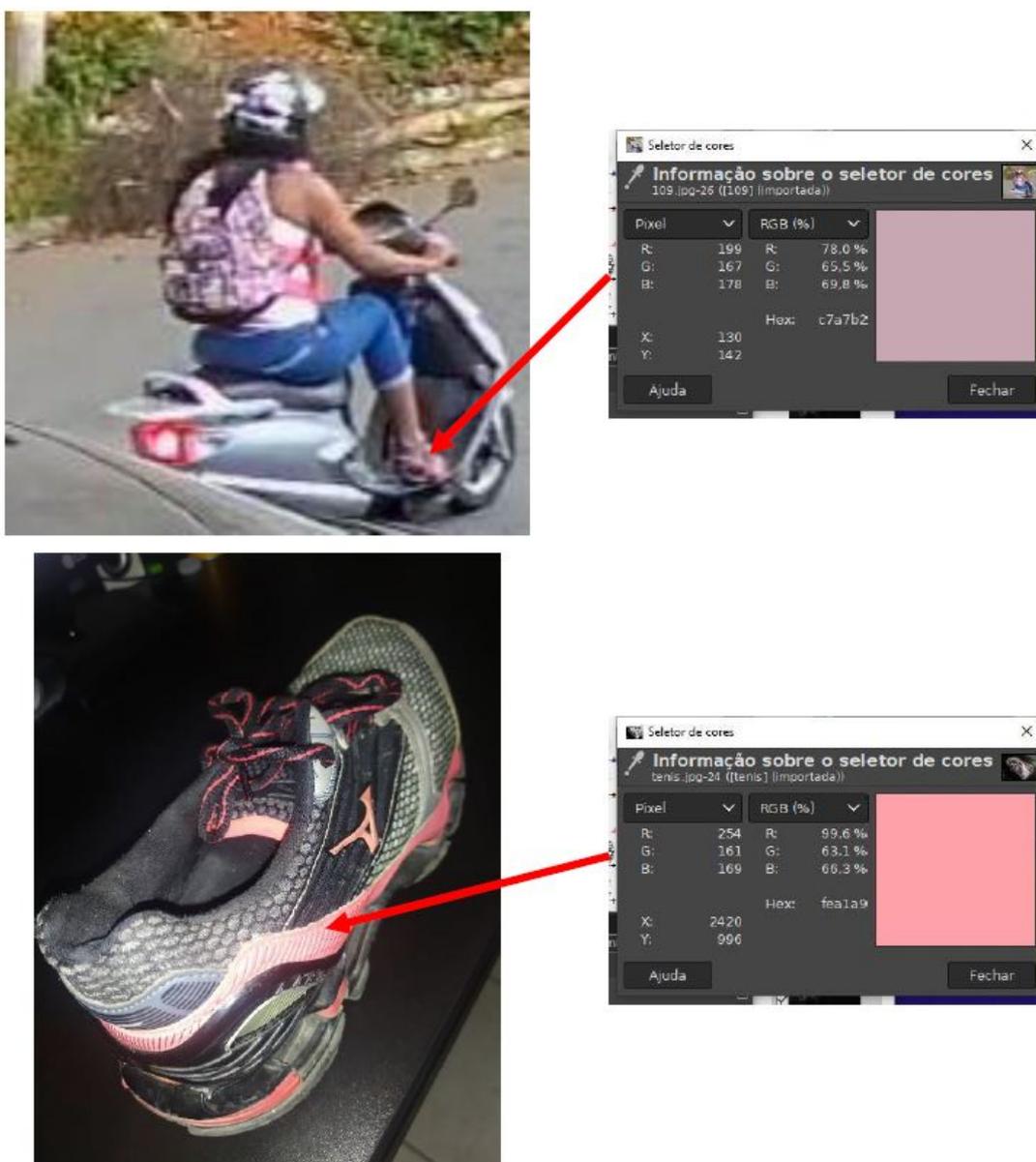
A calça examinada apresenta tonalidade de cor compatível com aquela da calça utilizada por um dos indivíduos presentes da ação delituosa. Fonte: produção própria adaptada de caso real. Descaracterizado para manutenção de sigilo.

Figura 2.17. – Detalhes da mochila examinada.



É possível encontrar na mochila examinada tonalidade de cor compatível com aquela da mochila utilizada por um dos indivíduos presentes da ação delituosa. Além disso, a mochila tem padrão de formações em mosaico, perceptualmente condizentes com a cor e os detalhes da mochila registrada na filmagem. Fonte: produção própria adaptada de caso real. Descaracterizado para manutenção de sigilo.

Figura 2.18. – Detalhes do tênis examinado.



É possível encontrar nos detalhes do tênis examinado tonalidade de cor compatível com aquela presente nos detalhes do tênis utilizada por um dos indivíduos presentes da ação delituosa. Fonte: produção própria adaptada de caso real. Descaracterizado para manutenção de sigilo.

Um dos quesitos apresentados pela autoridade policial foi “É possível afirmar que os objetos periciados aparecem nas imagens do delito, ou seja, são utilizados por algum dos autores?”.

Do ponto de vista técnico, esse quesito carece de consistência científica, posto que, mesmo que o conjunto de imagens tivesse a mais alta qualidade de captação e a comparação entre os elementos fosse “evidente” do ponto de vista leigo, ainda assim, seria impossível afirmar tratarem-se dos mesmos objetos apreendidos.

A atividade técnico-científica do Perito Criminal deve basear-se e limitar-se pelos saberes científicos.

Considerando-se que a interpretação de eventos únicos com base em informações reduzidas não tem subsídio técnico para a aplicação da estatística frequentista. Nesses casos, o Perito Criminal deve lançar mão do Teorema de Bayes e da Razão de Verossimilhança.

Uma possível argumentação é a que segue: diante das informações levantadas, é possível afirmar que há maior probabilidade de que os objetos periciados, a saber, a calça, a mochila e o tênis, sejam aqueles que aparecem nas imagens do delito e que as referidas peças de vestimenta e calçado tenham sido utilizadas por um dos indivíduos presentes na ação delituosa e uma menor probabilidade de que os referidos objetos sejam outros quaisquer que não aqueles que aparecem na filmagem.

Detalhando o raciocínio que leva à conclusão acima, empregando-se o arcabouço teórico e matemático detalhado à seção 2.3, pode-se delinear duas hipóteses antagônicas, a saber:

- H_P : as peças encaminhadas para comparação são as mesmas utilizadas por algum autor do crime e
- H_D : as peças encaminhadas para comparação são quaisquer outras que não aquelas utilizadas por algum autor do crime.

As informações levantadas foram:

- A calça examinada apresenta 5 (cinco) detalhes coincidentes àqueles da calça utilizada por um dos indivíduos presentes da ação delituosa;
- A calça examinada apresenta tonalidade de cor compatível com aquela da calça utilizada por um dos indivíduos presentes da ação delituosa;
- É possível encontrar na mochila examinada tonalidade de cor compatível com aquela da mochila utilizada por um dos indivíduos presentes da ação delituosa. Além disso, a mochila tem padrão de formações em mosaico, perceptualmente condizentes com a cor e os detalhes da mochila registrada na filmagem;
- É possível encontrar nos detalhes do tênis examinado tonalidade de cor compatível com aquela presente nos detalhes do tênis utilizada por um dos indivíduos presentes da ação delituosa.

Ausentes elementos numéricos para a produção do valor da Razão de Verossimilhança ao exemplo em tela – como por exemplo a prevalência numérica dessas características nos objetos usualmente comercializados na região em que houve a apreensão –, argumentativamente, é sabido *a priori* que há os mais variados padrões de cores, formatos e de desenhos para calças, calçados e mochilas. Acrescentando-se a isso o pressuposto *a priori* de que a apreensão dos elementos tenha sido em um ato único, e ainda baseando-se na comparação de diversos elementos concomitantemente presentes na calça, na mochila e no tênis examinados e seus correspondentes presentes na filmagem, é possível inferir que a hipótese H_P (de que as peças encaminhadas para comparação sejam as mesmas utilizadas por algum autor do crime) tenha mais força do que a hipótese H_D (de que as peças encaminhadas para comparação sejam quaisquer outras que não aquelas utilizadas por algum autor do crime), ou seja, pode-se estimar uma **Razão de Verossimilhanças** superior a 1 ($LR > 1$).

3. Identificação de falantes

Sumário

3.1. Som, voz, fala e abordagens forenses	86
3.1.1. Som e audição humana	86
3.1.2. A produção da voz	89
3.1.2.1. Articuladores Vocais	91
3.1.3. Reconhecimento de falantes	92
3.1.3.1. Fontes de Variabilidade no Reconhecimento de Falantes	94
3.1.3.2. Caracterização de Falantes	98
3.1.3.3. Abordagens na Identificação Forense de Falantes	99
3.1.3.4. Biometria	102
3.1.3.5. Disfarce da Voz	103
3.2. Processamento de sinais	103
3.2.1. Diferenças básicas entre sinais de tempo contínuo e de tempo discreto	105
3.2.2. Série de Fourier	108
3.2.3. Transformada de Fourier em Sinais de Tempo Contínuo	110
3.2.4. Transformada de Fourier em Sinais de Tempo Discreto	111
3.2.5. Teorema da Amostragem	112
3.2.6. Representações de sinais contínuos e discretos	113
3.2.7. Sistemas Digitais	114
3.2.8. Operador Convolução e Sistemas Lineares Invariantes no Tempo	115
3.3. Modelos para a voz	120

3.3.1. A escala mel	121
3.3.1.1. Coeficientes Cepstrais de Frequência Mel (MFCC)	122
3.3.1.2. Deltas e delta-deltas	123
3.3.1.3. Banco de filtros Mel	124
3.4. Análise de Componentes Principais	127
3.4.1. Obtenção dos componentes principais	128
3.5. SUVREL: Supervised Variational Relevance Learning	132

Nesta seção, tendo em mente o objetivo de produzir um texto o mais autocontido possível, apresento os principais conceitos relacionados à produção da fala, os modelos matemáticos da voz e o ferramental estatístico para processamento e análise de sinais. Também é introduzido o cabedal de verbetes empregados ao longo do presente trabalho. Tendo em vista que a literatura nacional sobre esse assunto ainda não é suficientemente ampla e tendo notado que muitos termos são encontrados apenas em inglês, procurarei traduzi-los e indicar suas formas originais como nota de rodapé. Manterei e empregarei diretamente no texto os termos no idioma original somente nos casos em que a tradução puder causar dúvida ou gerar interpretação dúbia.

3.1. Som, voz, fala e abordagens forenses

3.1.1. Som e audição humana

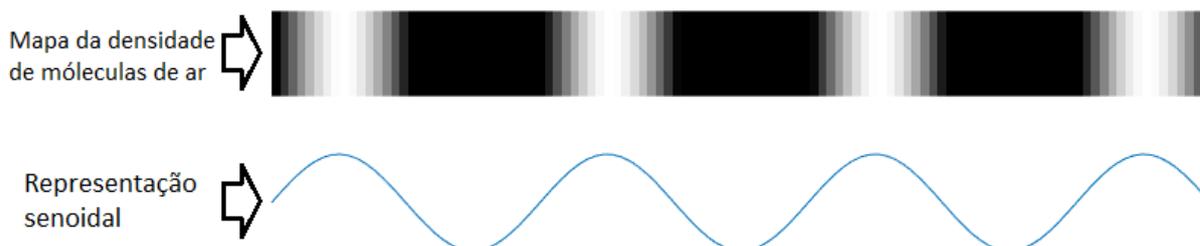
Som é uma onda longitudinal de pressão formada pela compressão e rarefação das moléculas de ar (fig. 3.1). A física do fenômeno de propagação das ondas sonoras em um gás envolve três relações características cíclicas (FEYNMAN; LEIGHTON; SANDS, 2010):

- o gás se desloca e altera de densidade;
- mudanças de densidades geram mudanças de pressão – “as variações de pressão e densidade são extremamente pequenas em relação aos valores de equi-

líbrio dessas grandezas, ou seja, a onda constitui uma pequena perturbação” (NUSSENZVEIG, 2002, p. 124);

- desigualdades de pressão geram deslocamento do gás – reiniciando o ciclo.

Figura 3.1. – Representação visual de uma frente de onda comparada à sua representação matemática.



Alto: uma fonte sonora provoca regiões de compressão e refração das moléculas de ar, portanto, uma onda longitudinal. Baixo: representação senoidal da mesma onda sonora. Fonte: produção própria. Inspirado em (HUANG; ACERO; HON, 2001, p. 21).

Usa-se comumente uma escala logarítmica de intensidades, posto que o estudo do som é fortemente fundamentado na forma como ocorre sua percepção e a sensibilidade auditiva se dá basicamente nessa escala.

O bel (B) é uma escala logarítmica de comparação relativa de intensidades. Uma diferença de 1 bel corresponde a uma relação de 10 em potência. Seu submúltiplo mais empregado é o decibel que corresponde a uma relação de potência de um décimo ($10^{\frac{1}{10}}$). Assim, é genericamente definida como:

$$1 \text{ dB} = 10 \log_{10} \left(\frac{B}{A} \right) \quad (3.1)$$

onde $B = 10^{\frac{1}{10}} A$.

O Nível de Pressão do Som (SPL³³) é uma medida absoluta em decibéis:

$$\text{dB}_{SPL} = 20 \log_{10} \left(\frac{P}{P_{ref}} \right) \quad (3.2)$$

onde a pressão de referência P_{ref} , correspondente ao limiar auditivo (0 dB_{SPL}), equi-

³³Do inglês, *Sound Pressure Level*.

vale a $P_{ref} = 2 \times 10^{-10}$ bar com frequência $f = 1$ kHz. Como exemplo, uma pressão $P = 10^3 \times P_{ref} = 2 \times 10^{-7}$ bar corresponde ao som de uma conversa entre interlocutores à distância de 1 m, cerca de 60 dB_{SPL} . Uma britadeira gera um som em torno de 120 dB_{SPL} . Na tabela 3.1, são listados alguns outros exemplos de níveis sonoros variados.

No presente trabalho, doravante, será padronizado o uso da escala absoluta dB_{SPL} que, por simplicidade, passarei a referir apenas como dB, salvo se essa escolha puder causar ambiguidade.

Tabela 3.1. – Níveis sonoros de fontes diversas.

Som	Nível dB_{SPL}
Limiar auditivo	0
Leve sussurro	10
Conversa suave	40
Conversa normal	60
Rua agitada	70
Guitarra acústica (a 30 cm de distância)	80
Tráfego de caminhões pesados	90
Metrô ouvido da plataforma	100
Ferramentas pesadas	110
Limiar de dor da audição humana	120
Pista de decolagem de aviões	130
<i>Sonic boom</i> ³⁴	140
Dano permanente à audição	150
Motor de jato (bem próximo)	160

Fonte: (HUANG; ACERO; HON, 2001, p. 22, tradução minha).

O limiar absoluto da audição é a quantidade máxima de energia sonora de um tom puro que não pode ser detectado por um ouvinte em um ambiente livre de ruídos. Trata-se de uma medida empírica que pode ser modelada aproximadamente pela equação 3.3 (HUANG; ACERO; HON, 2001, p. 23) cujo gráfico consta da figura

³⁴Som associado às ondas de choque produzidas por um objeto que viaje a velocidades superiores à velocidade do som.

3.2 em que as frequências estão representadas em escala logarítmica.

$$T(f) = 3,64 \left(\frac{f}{1000} \right)^{-0,8} - 6,5 \exp \left[-0,6 \left(\frac{f}{1000} - 3,3 \right)^2 \right] + 10^{-3} \left(\frac{f}{1000} \right)^4 \quad (3.3)$$

A abrangência do limiar absoluto da audição humana está entre 20 Hz e 20 kHz. Frequências abaixo de 20 Hz e acima de 20 kHz são designadas infrassons e ultrassons, respectivamente.

Figura 3.2. – Limiar Absoluto da Audição.

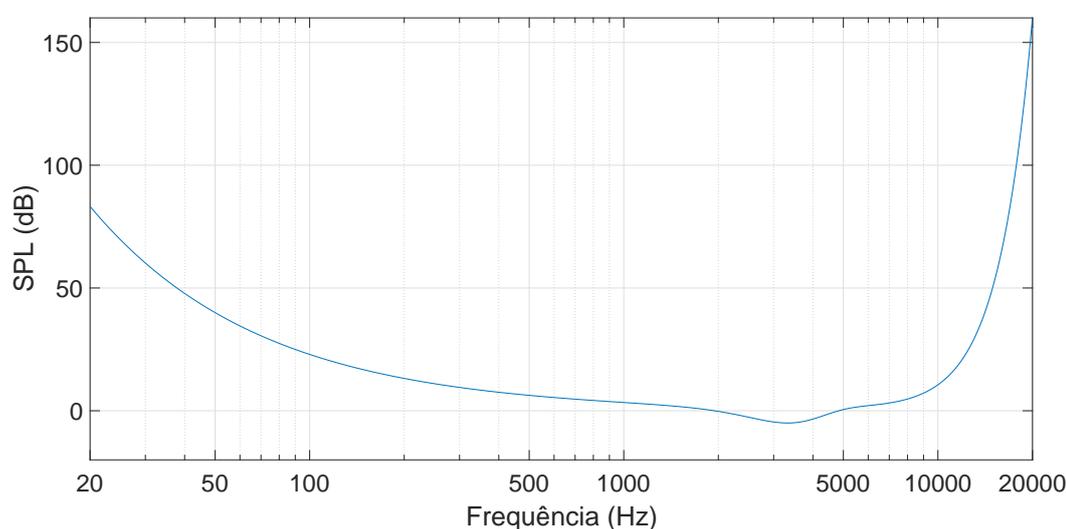


Gráfico do limiar absoluto da audição modelado pela equação 3.3. Fonte: produção própria.

3.1.2. A produção da voz

O estudo da produção da voz passou por diversas fases históricas e abordagens diferentes. Hipócrates (460-377 a.C.) já havia percebido que consoantes e vogais eram formadas de modos diferentes. Aristóteles (384-322 a.C.) comparava a laringe a uma flauta e a traqueia ao corpo de um instrumento. Leonardo da Vinci (1452-1519) contribuiu com suas pesquisas sobre a anatomofisiologia e a patologia da voz humana, além de ter realizado experimentos como o de apertar os pulmões de um ganso e observar que sons de tons variados eram produzidos. Nos séculos seguintes,

foram feitas várias descobertas sobre o papel dos pulmões, laringe, pregas vocais etc. na produção da voz e diversas teorias foram formuladas sendo a teoria mioelástica-aerodinâmica a mais aceita mundialmente.

A teoria mioelástica-aerodinâmica foi proposta em 1958 por Berg e ela combina a inter-relação de forças de duas naturezas: a da elasticidade dos músculos laríngeos e a das forças físicas aerodinâmicas da respiração. Nessa teoria, a produção de frequências graves e agudas é realizada pela contração dos músculos tensores, e a intensidade do som é resultado da contração dos músculos adutores, concomitantes às variações do fluxo aéreo. Essa teoria, desvincula frequência da intensidade e permite a análise desses dois parâmetros de forma individual.

Embora não tente expressar uma explicação para a produção da voz, sistemas complexos são bastante aplicados como método de análise (PETRY, 2002). Um sistema complexo deve apresentar certas características, tais como: ser um sistema não-linear; determinístico; imprevisível; com baixa dimensionalidade (poucas variáveis) e bastante sensível às alterações no sistema (ROSS; JAIN, 2007). Por essa metodologia, as pregas vocais e o fluxo aéreo transglótico representariam um sistema caótico (BEHLAU et al., 2001, p. 35–36).

Embora haja diversas teorias e o uso de equipamentos médicos cada vez mais modernos tenham permitido avanços surpreendentes no conhecimento da laringe, ainda não existe uma proposta abrangente que englobe toda a variedade da produção vocal humana normal e alterada pelas diversas disfunções possíveis (BEHLAU et al., 2001, p. 33–37).

Neste ponto, passo a revisar os tópicos básicos da produção da fala, tendo em vista que seu estudo influencia diretamente a codificação de sinais tendo (HUANG; ACERO; HON, 2001) como referência principal.

3.1.2.1. Articuladores Vocais

Na maioria dos idiomas, é possível classificar os sons produzidos pelo aparelho fonador em vogais e consoantes.

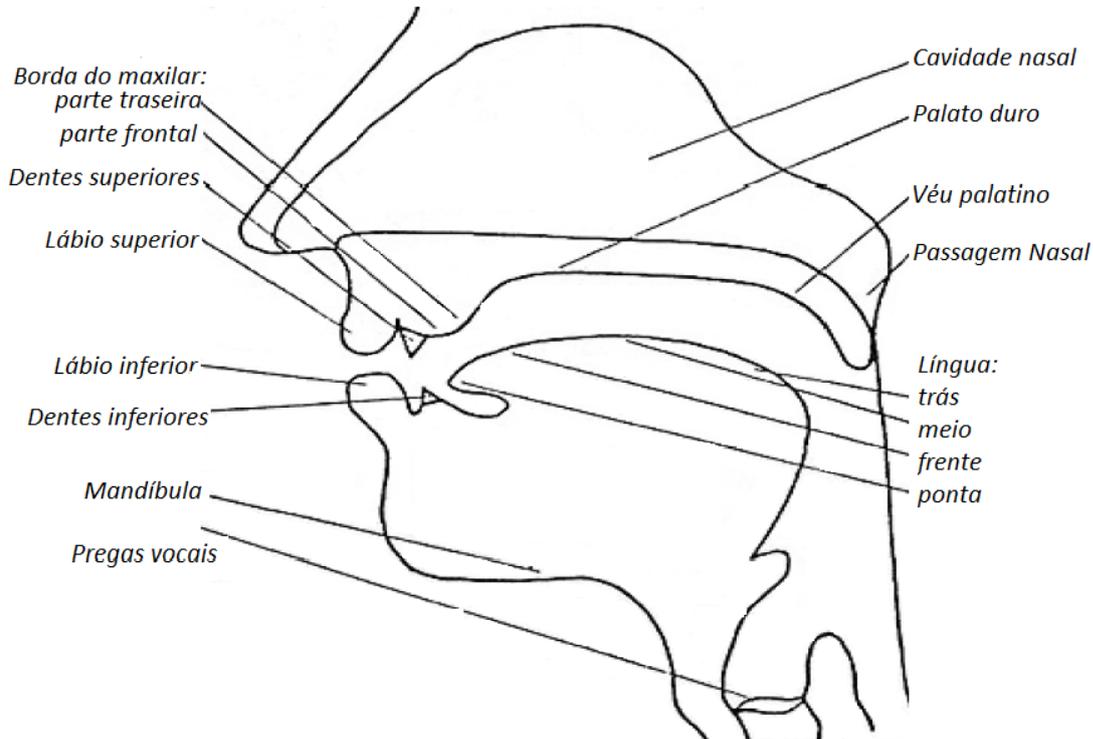
- consoantes são articuladas mediante constrictões na garganta e/ou obstruções na boca (língua, dentes, lábios)
- vogais são articuladas com poucas ou mínimas constrictões e obstruções.

Há várias propriedades relacionadas a diversos articuladores associados à anatomia individual do falante. Esses articuladores são bastante influenciados por um grande número de músculos que os posicionam e movimentam no trato fonador. Os principais componentes do trato fonador humano (fig. 3.3) são:

- Pulmões: fonte de ar durante a fala.
- Pregas vocais (laringe): produz sons *vozeados* (ou *sonoros*) quando as pregas vocais aproximam-se e mantêm-se juntas umas das outras em oscilações com vibrações periódicas (ex.: a palavra “Bato”). Quando as pregas vocais estão muito relaxadas para permitir vibrações periódicas, o som é dito *surdo* ou *não vozeado* (ex.: a palavra “Prato”).
- Véu palatino (ou palato mole): funciona como uma válvula que regula a passagem de ar através da cavidade nasal. Por exemplo, as letras *m* e *n* são produzidas com a aba do véu palatino aberto.
- Palato duro: juntamente com o palato mole, formam o comumente conhecido “céu da boca”. É uma superfície dura na qual a língua é pressionada para a articulação de consoantes.
- Língua: articulador flexível que fica afastado do palato na articulação de vogais.
- Dentes: empregado como ponto de apoio à língua para a produção de algumas consoantes.

- Lábios: elemento que pode se moldar para influenciar a qualidade de vogais e se fechar para impedir o fluxo de ar para algumas consoantes.

Figura 3.3. – Esquema do aparelho fonador humano.



Fonte: (HUANG; ACERO; HON, 2001, p. 41, tradução minha).

3.1.3. Reconhecimento de falantes

Uma característica humana natural é identificar vozes e relacioná-las a indivíduos. Essa habilidade é tão corriqueira que não é fácil perceber a complexidade e a diversidade de fatores envolvidos nesse processo. Além disso, de maneira natural, somos capazes de entender palavras em um idioma com grande habilidade para manter esse entendimento quando o idioma é falado em diferentes sotaques ou dialetos. Da mesma forma, com suficiente treinamento, uma pessoa é capaz de aprender novos idiomas. A habilidade de associar vozes a indivíduos é chamada de *Reconhecimento de Falantes* e a capacidade de identificar padrões de fala associando-os a palavras em um idioma é usualmente referido como *Reconhecimento de Voz*. O foco do presente

trabalho está no **Reconhecimento de Falantes**.

O reconhecimento de falantes pode ser feito de três maneiras (HANSEN; HASAN, 2015):

1. Reconhecimento natural de falantes³⁵: trata-se do reconhecimento de vozes familiares, que é característica inata ao ser humano, bastando que se ouça algumas vezes determinado falante sem que exista um treinamento consciente. Naturalmente, é possível identificar: prosódia (entonação, sotaque etc.), idioma, estilo de fala etc.
2. Reconhecimento Forense de Falantes: ouvintes treinados³⁶ comparam sistematicamente amostras de discursos e fornecem uma decisão fundamentada a respeito das similaridades dos falantes.
3. Reconhecimento Automático de Falantes: a análise completa do discurso e o processo completo de decisão é realizado usando análise computacional

O reconhecimento de falantes tem duas principais tarefas: a identificação de falantes e a verificação de falantes, que podem ser realizadas independentemente da maneira escolhida dentre as acima listadas.

A **Identificação de Falantes** tem por objetivo identificar um falante desconhecido em um conjunto de falantes conhecidos, ou seja, apontar o falante conhecido que soe mais próximo – ou que tenha os parâmetros objetivos mais semelhantes – ao falante desconhecido da amostra de áudio em análise. Em termos gerais, este método pode ser classificado como:

- cenário fechado³⁷: todos os falantes são conhecidos;
- cenário aberto³⁸: se o sujeito testado pode vir de fora de um grupo pré-definido. Nesse caso, são empregados os modelos mundial ou universal³⁹.

³⁵Usualmente referido na literatura por *Naïve Speaker Recognition*.

³⁶Normalmente, referido na literatura como *Auditory Skills*.

³⁷Também referido na literatura como *close scenario* ou *in-set scenario*.

³⁸Também referido na literatura como *open-set speaker recognition* ou *out-of-set speaker identification*.

³⁹*World Model* ou *Universal Background Model (UBM)*.

A identificação de falantes é muito usada para fins forenses.

Já a **Verificação de Falantes** é empregada quando um falante desconhecido é associado a uma identidade e o objetivo é verificar se isso é verdade. São comparadas duas amostras de voz ou elocução⁴⁰ e decide-se se são faladas pelo mesmo indivíduo. Isso é feito aplicando-se os modelos mundial e do possível falante⁴¹. A verificação de falantes é pouco usada para fins forenses.

3.1.3.1. Fontes de Variabilidade no Reconhecimento de Falantes

Diferentemente de impressões digitais, íris, geometria da mão, face etc., a fala humana é uma biometria de performance. A informação sobre a identidade do falante está embutida intrinsecamente no modo de falar e não necessariamente no conteúdo que é falado. Desse fato, decorre um enorme grau de variabilidade.

É fácil perceber que uma mesma pessoa não pronuncia as mesmas palavras do mesmo modo sempre, posto que ela pode variar seu estilo de fala⁴². Além disso, os equipamentos de gravação de áudio e os métodos de compactação para transmissão de áudios aumentam o problema. Assim, executar o reconhecimento de um falante é tarefa mais difícil quando a amostra provém de ligações telefônicas. Outros fatores que dificultam o reconhecimento de falantes são: doenças respiratórias, elevações do nível de estresse do falante, realização de outra tarefa simultânea ou falar com um diferente nível vocal – gritando ou cochichando.

Abaixo, são listadas algumas fontes de variabilidade no reconhecimento de falantes:

- Variabilidades originadas no falante:
 - Estresse situacional em uma tarefa: alterações cognitivas ou físicas.
 - Estilo/esforço vocal: alteração da fonação normal (cochichar, suavizar, “en-

⁴⁰ Do inglês, *utterance*.

⁴¹ Do inglês, *Claimed Speaker Model*.

⁴² Do inglês, *Style Shifting* ou *Intra-speaker Variability*.

grossar” a voz ou gritar); Efeito Lombard (HANSEN; VARADARAJAN, 2009): alteração do mecanismo de produção da fala para conversar efetivamente na presença de barulho; cantar versus falar.

- Emoção: comunicar-se enquanto vivencia-se um estado emocional (raiva, tristeza, alegria etc.).
- Fisiológica: doença, intoxicação ou efeitos da idade.
- Disfarce: o falante intencionalmente disfarça sua voz ou usa um sistema de conversão de voz.
- Variabilidades originadas na conversação:
 - Humano para humano: idioma ou dialeto falado; se a fala é lida, espontânea, em uma conversa ou disfarçada; monólogo, conversa de duas vias, discurso para um público ou TV, rádio, grupo de discussão.
 - Humano para máquina: *input* de voz para um computador⁴³; interações com sistemas de voz (telefone, sistemas de diálogo etc.).
- Variabilidades baseadas na tecnologia (ou, genericamente, variabilidades externas):
 - Eletromecânica: canais de transmissão, microfones etc.
 - Ambiental: ruído de fundo, acústica ambiental, reverberação, distância do microfone etc.
 - Qualidade dos dados: duração, taxa de amostragem⁴⁴, qualidade de gravação e codec/compressão.

⁴³Do inglês, *prompted-speech*.

⁴⁴Do inglês, *sampling rate*.

3.1.3.1.1 Parâmetros Acústicos

Como resultado da extensa quantidade de fontes de variabilidade intra e interlocutores, o reconhecimento forense de falantes lança mão de diversos parâmetros acústicos. Abaixo, reproduzo os parâmetros elencados por Sanches (2012):

1. Voz

- Aspectos Funcionais

- Qualidade vocal: produto final da voz. Conjunto de características acústicas provenientes da onda sonora.
- Estabilidade vocal ou fonatória
- Ressonância e Foco: amplificação do som emitido pelas pregas vocais por meio das cavidades: laríngea, faríngea, oral e bucal.
- Prosódia: descreve as propriedades acústicas da fala que não podem ser previstas pela transcrição ortográfica. Trata da acentuação tônica das palavras.
- Ritmo: o ritmo e a velocidade de fala dizem respeito à agilidade de encadear os diferentes ajustes motores necessários para a fala. Eles fazem parte da Prosódia da fala.
- Melodia / Modulação
- Uso de pausas
- Uso de ênfases

- Aspectos Acústicos

- Intensidade
- Jitter: indica a variabilidade da frequência fundamental a curto prazo, medida entre ciclos glóticos vizinhos.

- Shimmer: representa as alterações irregulares na amplitude dos ciclos, de um ciclo ao outro.
- Frequência modal: registro acústico da frequência que mais é utilizada na fala habitual.
- Quebra de frequência: registro acústico da frequência que mais é utilizada na fala habitual.
- Quebra de sonoridade: ruptura momentânea de emissão sonora durante a produção do som.
- Gama tonal: faixa de notas acima e abaixo da frequência habitualmente utilizada durante um discurso.
- Projeção / harmônicos: Componente inteiro e múltiplo de uma Frequência Fundamental. Quanto mais harmônico existir, melhor é a projeção da voz.
- Tremor
- Formantes

2. Fala

- Articulação: ajustes motores dos órgãos fonoarticulatórios na produção e formação dos sons, e ao encadeamento destes na fala.
- Pronúncia: refere-se ao uso de determinadas substituições de sons nas palavras ou variações articulatórias de um mesmo som, e é resultado de um condicionamento fonológico resultante da exposição ao código linguístico de uma população em particular.
- Velocidade: o ritmo e a velocidade de fala dizem respeito à agilidade de encadear os diferentes ajustes motores necessários para a fala. Eles fazem parte da prosódia da fala.

- Desvios: qualquer alteração identificada na Voz, Fala e Linguagem.

3. Linguagem

- Competência comunicativa
- Conexidade
- Sotaque
- Regionalismo: variações faladas por comunidades geograficamente definidas.
- Socioleto: variações faladas por comunidades socialmente definidas.
- Idioleto: variação individual correlacionada com experiências pessoais, ambiente, formação e personalidade.

4. Respiração

- Tipo
- Modo
- CPFA: coordenação realizada entre a respiração, voz e fala.
- Uso de ar de reserva

3.1.3.2. Caracterização de Falantes

Cada falante tem alguns traços característicos em sua voz que são únicos, como a fisiologia de seu trato vocal e os hábitos adquiridos na articulação das palavras. Duas pessoas distintas podem ter algum ou alguns parâmetros característicos idênticos.

Um parâmetro característico deve ter as seguintes propriedades ideais (NOLAN, 1983):

1. Mostrar alta variabilidade entre falantes e pouca variabilidade para o mesmo falante
2. Ser resistente a tentativas de disfarce ou mimetismo

3. Ter uma alta frequência de ocorrência em materiais relevantes
4. Ser robusto na transmissão
5. Ser relativamente fácil de extrair ou medir

Os parâmetros característicos podem ser classificados como:

- Auditivos / Acústicos
 - Auditivos: podem ser objetivamente descritos quando ouvidos
 - Acústicos: parâmetros matematicamente definidos extraídos por algoritmos automáticos
- Linguísticos / Não-linguísticos
 - Linguísticos: dentro da estrutura de uma língua, línguas ou dialetos
 - Não-linguísticos: qualidade da fala (nasalizada, áspera etc), fluência, pausas, ritmo, frequência fundamental média e sons não-verbais (tosses, risadas etc.)
- De curto prazo/ De longo prazo
 - Curto prazo: comparação direta de difones, sílabas, fonemas (e.g.: o grafema “r”) e transição consoante-vogal
 - Longo prazo: geralmente são médias dos parâmetros de curto-prazo (e.g.: *pitch* e contornos dos formantes). Se o parâmetro é extraído de uma conversa completa, é comumente referido por parâmetro do nível de elocução⁴⁵.

3.1.3.3. Abordagens na Identificação Forense de Falantes

A perícia criminal é uma das ferramentas empregadas pela Justiça para auxiliar a formação da convicção dos Magistrados em busca da verdade real. O perito criminal trabalha em busca do esclarecimento, à luz do conhecimento científico, de evidências

⁴⁵Do inglês, *utterance-level feature* ou *utterance feature*.

levantadas e relacionadas a crimes e ele pode ter diversas formações específicas tais como biologia, química, ciências da computação, farmácia, engenharia, física, fonética, fonoaudiologia etc.

A física forense engloba técnicas úteis ao esclarecimento das causas de acidentes automobilísticos, navais, aéreos etc., à perícia de locais de crimes, desastres, além de embasar o desenvolvimento de novas tecnologias capazes de otimizar a análise de evidências diversas. Em termos jurídicos, a Física Forense é “a parte da Física destinada à observação, análise e interpretação dos fenômenos físicos naturais de interesse judiciário”.

Há diversas abordagens atualmente empregadas simultaneamente no campo da identificação forense de falantes. As três principais são:

- Abordagem auditiva: baseada na experiência e treinamento de foneticistas e fonoaudiólogos, é totalmente subjetiva. O resultado é apresentado em uma declaração formal em que são descritas as bases das evidências e das conclusões;
- Abordagem auditivo-espectrográfica: baseia-se na comparação visual entre espectrogramas. É bastante empregada porque facilita o entendimento por parte dos jurados, mas tem o porém de ser subjetiva e depender muito da experiência do perito examinador;
- Abordagem acústico-fonética: feita por peritos treinados em fonética e acústica, requer medidas acústicas quantitativas e análise estatística dos resultados. Apesar de ser mais objetiva, também tem elementos subjetivos, como exemplo, um perito pode identificar que a fala soa afetada por estresse e usar isso para fazer a análise objetiva, mas a conclusão sobre o estresse terá sido subjetiva.

Percebe-se que a palavra final em um trabalho de comparação de falantes ou locutores é dada por um perito foneticista ou fonoaudiólogo que, apesar de usar recursos computacionais para facilitar seu trabalho, depende principalmente de sua habilidade auditiva (CHARTRAND; PERETZ; BELIN, 2008). Assim, a habilidade de

ouvir tecnicamente desempenha papel fundamental na análise forense e isso decorre da habilidade natural humana em identificar vozes e inferir sentimentos e/ou emoções de seu interlocutor. São pontos fortes dos ouvintes humanos:

- identificar falantes com grande precisão, até em condições difíceis;
- identificar facilmente idiossincrasias: se um falante tosse de uma determinada maneira isto pode representar algo no discurso.

São pontos fracos dos ouvintes humanos:

- São mais suscetíveis a tendenciosidade contextual (e.g.: se um suspeito confessou o crime, o perito tende a identificar positivamente a voz do suspeito com a amostra);
- Sujeito a erros;
- Não podem lembrar de uma voz por muito tempo.
- Mesmo para vozes familiares, o ouvinte pode ter a sensação de reconhecer a voz, mas não conseguir identificar de quem é;
- Ouvintes não treinados não conseguem distinguir sotaques que diferem pouco entre si;
- A atenção cai após algum tempo de trabalho;
- Os resultados expressos em *LR* podem não ser consistentes entre peritos diferentes;
- Ouvintes humanos (mesmo peritos) tendem a identificar alguém em uma gravação se eles esperam encontrar determinado sujeito nela.

Cabe notar que apesar de ser necessária uma habilidade especial do perito, não deixa de causar estranheza que não haja diversos programas que façam essa tarefa de forma automatizada. Essa contradição ao senso comum é reconhecida na comunidade de peritos em áudio como “efeito CSI” referindo-se à série televisiva de ficção *Crime Scene Investigation*. Nos Estados Unidos, nota-se que juízes e membros de júri

familiarizados com essa série chegam à corte com expectativas sobre as capacidades de melhorias de sinais de áudio e identificação positiva de voz totalmente incompatíveis com a realidade (MAHER, 2009, p. 93).

3.1.3.4. Biometria

Biometria é o ramo da ciência que busca estabelecer a identidade de um indivíduo baseando-se em traços físicos ou comportamentais associados a uma pessoa. Há vários sistemas de identificação biométricas tais como impressão digital, íris, face, geometria da mão, impressão da palma, estrutura de veias no dedo, assinatura, voz, padrão de digitação em teclado etc. cada qual com seu grau de eficiência, aplicabilidade e restrições.

Um sistema biométrico consiste em capturar um traço específico de um indivíduo e compará-lo com um traço biométrico já armazenado em um banco de dados (ROSS; JAIN, 2007) para determinar a identidade de uma pessoa (identificação) ou validar sua identidade (verificação). A aplicabilidade dessa técnica engloba programas de segurança de ponta, investigações criminais, pontos de acesso lógico (*logins* em computadores) e equipamentos de vigilância. Por conta dessa gama de utilizações, o reconhecimento biométrico é considerado um grande desafio para diversas áreas científicas, pois requer alto nível de eficiência para sua aceitação.

A metodologia estatística convencional empregada para mensurar o nível de confiabilidade da biometria é medida em termos de:

1. Níveis de falsa aceitação: quando o acesso é dado a uma pessoa não autorizada;
2. Níveis de falsa rejeição: quando o verdadeiro usuário é erroneamente bloqueado;
3. Taxa de erro igual (EER⁴⁶): quando o nível percentual de falsas aceitações é igual ao percentual de falsas rejeições.

⁴⁶Do inglês, *equal error rate*.

3.1.3.5. Disfarce da Voz

Sabe-se que há pessoas capazes de imitar vozes com elevado grau de maestria. Não só imitar, como também disfarçar sua voz fazendo uso de objetos que travem sua articulação maxilar (canetas), panos sobre a boca ou ainda aumentando ou diminuindo a frequência e a velocidade de sua fala.

Em adendo ao bojo de tópicos que compõem essa introdução, breve pesquisa foi feita a esse respeito. Em especial, o trabalho de Mestrado em Linguística de Gillier (2011), chamou a atenção por apresentar diversos parâmetros de classificação dos disfarces de voz e por estabelecer elementos de comparação objetivos que podem vir a ser empregados em um sistema computacional.

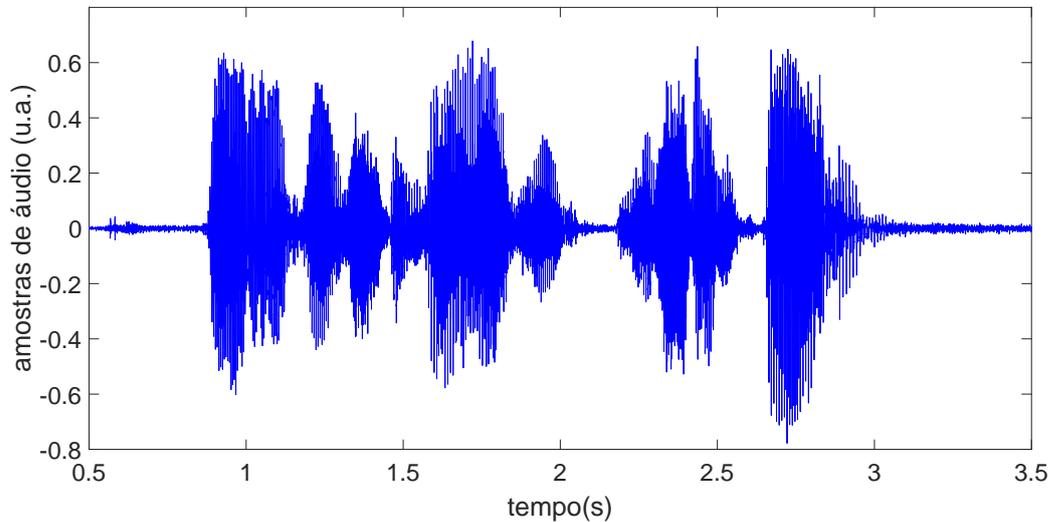
3.2. Processamento de sinais

Nesta etapa, faço uma revisão dos elementos de Processamento de Sinais necessários ao desenvolvimento deste trabalho. Utilizarei como principal fonte o texto de Oppenheim, Willsky e Nawab (2010), por se tratar de livro de referência no tema, além de recorrer ao trabalho de Huang, Acero e Hon (2001, cap. 5).

Genericamente, sinais são registros sistemáticos da variação de uma ou mais medidas ao longo do tempo (ex.: som) ou espaço (ex.: imagens) e são representados matematicamente como funções de uma ou mais variáveis independentes. Um sinal de voz é graficamente representado pela pressão acústica em função do tempo (fig. 3.4). Uma imagem (em escala de cinzas) pode ser entendida como a medida de brilho em função de duas variáveis espaciais. Como o foco deste trabalho são sinais sonoros, a variável independente sempre será o tempo.

Os sinais podem ser classificados entre contínuos e discretos. Para denotar essa diferença, é usual empregar a letra t , entre parênteses, para a variável independente de tempo contínuo. Já para a variável independente de tempo discreto, utiliza-se n entre

Figura 3.4. – Exemplo de série temporal de um sinal de voz.



No exemplo, foi registrada a frase “Universidade de São Paulo”. Fonte: produção própria.

colchetes (fig. 3.5).

Sinais contínuos e discretos podem ser analisados basicamente pelas mesmas ferramentas matemáticas, mas, como sinais de tempo discreto somente são definidos para valores inteiros da variável independente, surgem algumas diferenças que evidenciarei nas seções seguintes.

A primeira etapa do processamento de sinais é converter o sinal de entrada analógico $x(t)$ em um sinal digital essencialmente discreto $x[n]$ formado pela amostragem periódica, a intervalos regulares T , do sinal analógico $x(t)$. Portanto:

$$x[n] = x(nT) \quad (3.4)$$

onde T é o período de amostragem. Com base nessa definição, a frequência de amostragem é $F_S = \frac{1}{T}$.

A escolha da frequência de amostragem é item fundamental na análise de sinais, tendo em vista o que dispõe o Teorema da Amostragem (sec. 3.2.5) que, de modo simplificado, explica que um sinal analógico $x(t)$ somente pode ser univocamente

Figura 3.5. – Sinais de tempo discreto e de tempo contínuo.

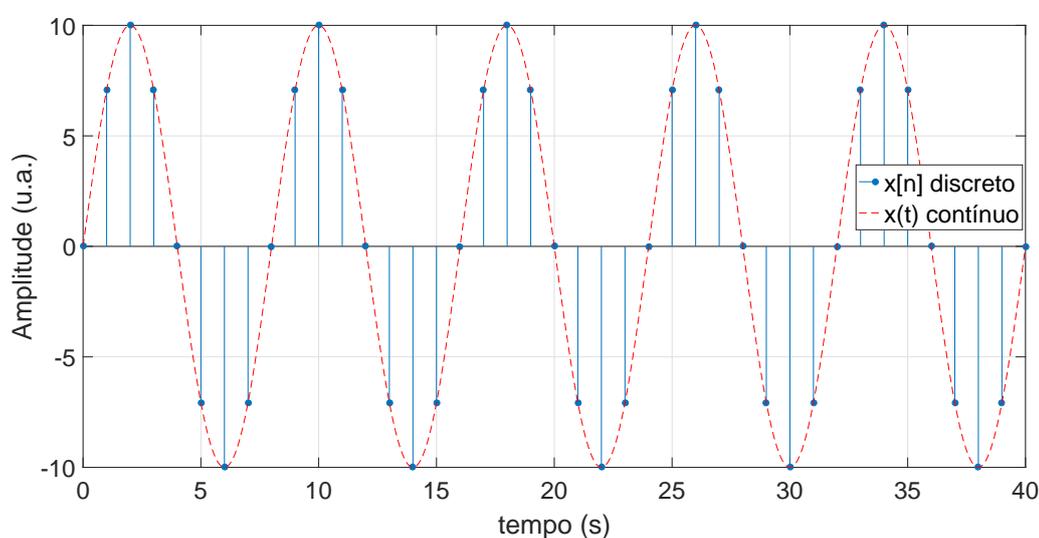


Gráfico da função $y = 10 \sin \frac{\pi x}{4}$ tomada discretamente em $x[n]$ com $n = 0, 1, \dots, 40$ e representada com $x(t)$ contínuo. Fonte: produção própria.

recuperado a partir de $x[n]$ se o sinal analógico não tiver energia nas frequências superiores à *frequência de Nyquist*, definida como $\frac{F_S}{2}$.

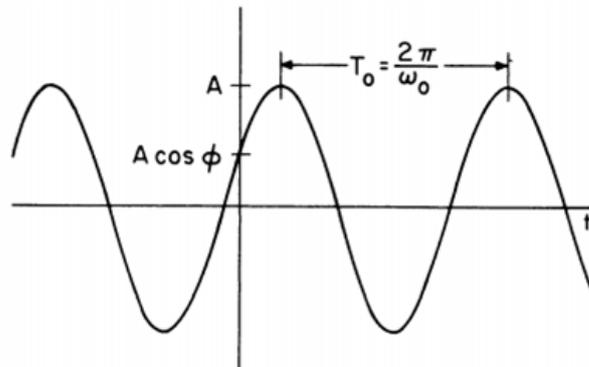
3.2.1. Diferenças básicas entre sinais de tempo contínuo e de tempo discreto

A principal diferença entre um sinal de tempo contínuo e um de tempo discreto é que este somente é definido para valores inteiros da variável independente. Isso pode ser demonstrado pelo seguinte exemplo extraído de (OPPENHEIM; WILLSKY; NAWAB, 2010).

Tome-se um sinal senoidal contínuo da forma $x(t) = A \cos(\omega_0 t + \phi)$ (fig. 3.6). Ele será periódico se $x(t) = x(t + T_0)$, onde T_0 é o período (o menor valor de $T_0 \stackrel{\text{def}}{=} \text{período}$), e se:

$$A \cos(\omega_0 t + \phi) = A \cos(\omega_0 t + \underbrace{\omega_0 T_0}_{2\pi m} + \phi) \quad (3.5)$$

Figura 3.6. – Gráfico do sinal contínuo $x(t) = A \cos(\omega_0 t + \phi)$.



Fonte: (OPPENHEIM; WILLSKY; NAWAB, 2010, cap. 1).

Portanto, $T_0 = \frac{2\pi m}{\omega_0}$ é condição para definir a periodicidade de um sinal contínuo. Porém, como m pode assumir qualquer valor neste tipo de sinal, a condição suficiente é que $T_0 = \frac{2\pi}{\omega_0}$. Assim, um deslocamento no tempo provoca uma mudança de fase e vice-versa:

Deslocamento no tempo \Leftrightarrow Mudança de Fase

$$A \cos[\omega_0(t + t_0)] = A \cos[\omega_0 t + \underbrace{\omega_0 t_0}_{\phi_{novo}}] \quad (3.6)$$

$$A \cos[\omega_0(t + t_0) + \phi] = A \cos[\omega_0 t + \underbrace{\omega_0 t_0 + \phi}_{\phi_{novo}}] \quad (3.7)$$

Já no caso de um sinal senoidal discreto da forma $x[n] = A \cos(\Omega_0 n + \phi)$, um deslocamento no tempo gera uma mudança de fase:

$$A \cos[\Omega_0(n + n_0)] = A \cos[\Omega_0 n + \underbrace{\Omega_0 n_0}_{\phi_{novo}}] \quad (3.8)$$

mas, a recíproca não é sempre válida. Definindo-se um período N , analogamente ao caso de sinais contínuos, tal que $x[n] = x[n + N]$ (o menor inteiro $N \stackrel{\text{def}}{=} \text{período}$), então:

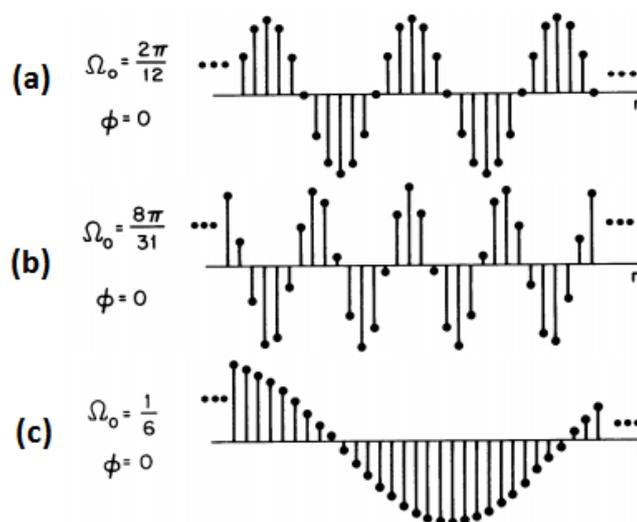
$$A \cos[\Omega_0(n + N) + \phi] = A \cos[\Omega_0 n + \underbrace{\Omega_0 N}_{\text{inteiro múltiplo de } 2\pi} + \phi] \quad (3.9)$$

Portanto, a condição de periodicidade de um sinal discreto é:

$$\Omega_0 N = 2\pi m \implies N = \frac{2\pi m}{\Omega_0} \quad (3.10)$$

onde N, m devem ser inteiros e o menor N (se existir) será o período. Alguns exemplos ilustrativos da figura 3.7: (a) $\Omega_0 = \frac{2\pi}{12}$ gera um sinal discreto periódico com $m = 1$ e $N = 12$; (b) $\Omega_0 = \frac{8\pi}{31}$ também é periódico com $m = 4$ e $N = 31$ (m e N não têm fatores comuns entre si, logo, $N = 31$ é o menor inteiro possível) e (c) embora a inspeção visual sugira um sinal periódico, com $\Omega_0 = \frac{1}{6}$, impõe-se um $N = 12\pi$ não inteiro, portanto, este sinal discreto é aperiódico.

Figura 3.7. – Gráficos do sinal discreto $x[n] = A \cos(\Omega_0 n + \phi)$.



Exemplos periódicos: (a) $\Omega_0 = \frac{2\pi}{12}$ com $m = 1$ e $N = 12$ e (b) $\Omega_0 = \frac{8\pi}{31}$ com $m = 4$ e $N = 31$. Exemplo aperiódico: (c) $\Omega_0 = \frac{1}{6}$, impõe um $N = 12\pi$ não inteiro. Fonte: (OPPENHEIM; WILLSKY; NAWAB, 2010, cap. 1).

Essas diferenças implicam em transformadas de Fourier com características distintas para sinais contínuos e discretos.

Tabela 3.2. – Comparação entre sinais de tempo contínuo $x(t)$ e de tempo discreto $x[n]$.

$x(t) = A \cos(\omega_0 t + \phi)$	$x[n] = A \cos(\Omega_0 n + \phi)$
Sinais distintos para valores distintos de ω_0	Sinais idênticos para valores de Ω_0 separados por múltiplos de 2π
Periódico para qualquer escolha de ω_0	Periódico somente se $\Omega_0 = \frac{2\pi m}{N}$ e alguns valores de $N > 0$ e m
Frequência fundamental ω_0	Frequência fundamental $\frac{\Omega_0}{m}$

Fonte: Adaptado de (OPPENHEIM; WILLISKY; NAWAB, 2010, cap. 1).

3.2.2. Série de Fourier

Seja um sinal periódico da forma

$$x(t) = \exp\{j\omega_0 t\} \quad (3.11)$$

de frequência fundamental ω_0 e período fundamental $T = \frac{2\pi}{\omega_0}$. Pelas propriedades dos números complexos pode-se associar a este sinal a ideia de uma exponencial complexa harmônica definida por:

$$\phi_k(t) = \exp\{jk\omega_0 t\} = \exp\left\{jk\left(\frac{2\pi}{T}\right)t\right\} \quad (3.12)$$

onde $k = 0, \pm 1, \pm 2, \dots$.

Sinais periódicos da equação 3.11 construídos com os harmônicos da equação 3.12 têm frequências fundamentais que são múltiplas de ω_0 e, assim, também periódicas com período T .

A série de Fourier (OPPENHEIM; WILLISKY; NAWAB, 2010, cap. 3) é a combinação linear de exponenciais complexas harmonicamente relacionadas na forma:

$$x(t) = \sum_{k=-\infty}^{+\infty} a_k \exp\{jk\omega_0 t\} = \sum_{k=-\infty}^{+\infty} a_k \exp\left\{jk\left(\frac{2\pi}{T}\right)t\right\} \quad (3.13)$$

Os coeficientes de 3.13 são calculados por:

$$a_k = \frac{1}{T} \int_T x(t) \exp\{-jk\omega_0 t\} dt = \frac{1}{T} \int_T x(t) \exp\left\{-jk \left(\frac{2\pi}{T}\right) t\right\} dt \quad (3.14)$$

As equações 3.13 e 3.14 são denominadas equações de síntese e de análise, respectivamente.

Em consequência dos motivos explicitados na seção 3.2.1, o par de equações da série de Fourier de tempo discreto é dado por:

$$x[n] = \sum_{k=\langle N \rangle} a_k \exp\{jk\Omega_0 n\} = \sum_{k=\langle N \rangle} a_k \exp\left\{jk \left(\frac{2\pi}{N}\right) n\right\} \quad (3.15)$$

e

$$a_k = \frac{1}{N} \sum_{n=\langle N \rangle} x[n] \exp\{-jk\Omega_0 n\} = \frac{1}{N} \sum_{n=\langle N \rangle} x[n] \exp\left\{-jk \left(\frac{2\pi}{N}\right) n\right\} \quad (3.16)$$

em que o intervalo do somatório compreende os valores de n em apenas um período N .

Mais explicitamente, como a condição de periodicidade de um sinal discreto é $N = \frac{2\pi m}{\Omega_0}$ e como exponenciais complexas que diferem em frequência por um múltiplo de 2π são iguais, então:

$$\phi_k[n] = \phi_{k+rN}[n] \quad (3.17)$$

com r sendo um inteiro qualquer.

Portanto, escrevendo $x[n]$ em termos de k variando de 0 a $N-1$ e de 1 a N , observa-se que:

$$x[n] = a_0\phi_0[n] + a_1\phi_1[n] + a_2\phi_2[n] + \cdots + a_{N-1}\phi_{N-1}[n] \quad (3.18)$$

$$x[n] = a_1\phi_1[n] + a_2\phi_2[n] + \cdots + a_{N-1}\phi_{N-1}[n] + a_N\phi_N[n] \quad (3.19)$$

logo:

$$\phi_0[n] = \phi_N[n] \Rightarrow a_0 = a_n \Rightarrow a_k = a_{k+N} \quad (3.20)$$

Assim, quando são considerados mais do que N valores sequenciais de k , os valores de a_k se repetem periodicamente com período N .

3.2.3. Transformada de Fourier em Sinais de Tempo Contínuo

Tomando-se um um sinal contínuo $x(t)$, sua transformada de Fourier $\mathcal{F}(x(t)) = X(\omega)$ é definida através das equações de análise (eq. 3.21) e de síntese (eq. 3.22):

$$X(\omega) = \int_{-\infty}^{+\infty} x(t) \exp\{-j\omega t\} dt \quad (3.21)$$

$$x(t) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} X(\omega) \exp\{j\omega t\} d\omega \quad (3.22)$$

A conversão de um sinal entre os espaços vetoriais do tempo e de frequências (e vice-versa) é usualmente representado por $x(t) \xleftrightarrow{\mathcal{F}} X(\omega)$. A transformada de Fourier, em termos de seus componentes complexos, é:

$$X(\omega) = \text{Re}\{X(\omega)\} + j \text{Im}\{X(\omega)\} = |X(\omega)| \exp\{j \angle X(\omega)\} \quad (3.23)$$

Algumas propriedades da transformada de Fourier em tempo contínuo são:

- Simetria:

$$\text{Se } x(t) \text{ for real} \Rightarrow X(-\omega) = X^*(\omega) \quad (3.24)$$

$$\text{Se } x(t) \text{ for par} \begin{cases} \text{Re}\{X(\omega)\} & = \text{Re}\{X(-\omega)\} \\ |X(\omega)| & = |X(-\omega)| \end{cases} \quad (3.25)$$

$$\text{Se } x(t) \text{ for ímpar} \begin{cases} \text{Im}\{X(\omega)\} & = -\text{Im}\{X(-\omega)\} \\ \angle(X\omega) & = -\angle X(-\omega) \end{cases} \quad (3.26)$$

- Relação de Parseval:

$$\int_{-\infty}^{+\infty} |x(t)|^2 dt = \frac{1}{2\pi} \int_{-\infty}^{+\infty} |X(\omega)|^2 d\omega \quad (3.27)$$

- Deslocamento no tempo:

$$x(t - t_0) \xleftrightarrow{\mathcal{F}} \exp\{-j\omega t_0\} X(\omega) \quad (3.28)$$

- Diferenciação:

$$\frac{dx(t)}{dt} \xleftrightarrow{\mathcal{F}} j\omega X(\omega) \quad (3.29)$$

- Integração:

$$\int_{-\infty}^t x(\tau) d\tau \xleftrightarrow{\mathcal{F}} \frac{1}{j\omega} X(\omega) + \pi X(0) \delta(\omega) \quad (3.30)$$

- Linearidade:

$$ax_1(t) + bx_2(t) \xleftrightarrow{\mathcal{F}} aX_1(\omega) + bX_2(\omega) \quad (3.31)$$

- Convolução (detalhada à sec. 3.2.8):

$$h(t) * x(t) \xleftrightarrow{\mathcal{F}} H(\omega) X(\omega) \quad (3.32)$$

- Modulação

$$h(t)x(t) \xleftrightarrow{\mathcal{F}} \frac{1}{2\pi} [H(\omega) * X(\omega)] \quad (3.33)$$

3.2.4. Transformada de Fourier em Sinais de Tempo Discreto

Tendo em vista as diferenças elencadas na seção 3.2.1, a transformada de Fourier de um sinal de tempo discreto $\mathcal{F}\{x[n]\} = X(\Omega)$ é definida através das equações

de análise (eq. 3.34) e de síntese (eq. 3.35):

$$X(\Omega) = \sum_{n=-\infty}^{+\infty} x[n] \exp\{-j\Omega n\} \quad (3.34)$$

$$x[n] = \frac{1}{2\pi} \int_{2\pi} X(\Omega) \exp\{j\Omega n\} d\Omega \quad (3.35)$$

em que se observa que a equação de síntese deixa de ser definida de $-\infty$ a $+\infty$ e passa a ser definida para um período 2π .

Algumas das propriedades da transformada de Fourier em tempo discreto, usualmente representada por $x[n] \xleftrightarrow{\mathcal{F}} X(\Omega)$, são análogas àsquelas de tempo contínuo: simetria, deslocamento no tempo e linearidade (sec. 3.2.3). São propriedades distintas:

- Periodicidade:

$$X(\Omega) = X(\Omega + 2\pi m) \quad (3.36)$$

- Modulação:

$$x_1[n]x_2[n] \xleftrightarrow{\mathcal{F}} \frac{1}{2\pi} \int_{2\pi} \underbrace{X_1(\theta)X_2(\Omega - \theta)d\theta}_{\text{convolução periódica}} \quad (3.37)$$

Cabe frisar, que a modulação em tempo contínuo é definida por:

$$x_1(t)x_2(t) \xleftrightarrow{\mathcal{F}} \frac{1}{2\pi} \int_{-\infty}^{+\infty} \underbrace{X_1(\rho)X_2(\omega - \rho)d\rho}_{\text{convolução aperiódica}} \quad (3.38)$$

3.2.5. Teorema da Amostragem

Dadas amostras igualmente espaçadas de $x(t)$:

$$x(nT), \text{ onde } n = 0, \pm 1, \pm 2, \dots \quad (3.39)$$

e seja $x(t)$ de frequência limitada, ou seja, existe uma frequência de corte ω_0 , tal que $X(\omega) = 0$ para $|\omega| > \omega_0$.

Se a frequência de amostragem ω_s (definida como $\omega_s \stackrel{\text{def}}{=} \frac{2\pi}{T}$), obedecer a relação:

$$\omega_s > 2\omega_m \quad (3.40)$$

então, $x(t)$ é univocamente recuperada com um filtro passa-baixa.

Quando a frequência de amostragem não é escolhida de modo a obedecer a equação 3.40, observa-se o efeito de rebatimento ou recobrimento (*aliasing*), ou distorção entre o sinal original e o sinal recuperado via interpolação. A figura 3.8 exemplifica esse resultado.

A tabela 3.3, ilustra algumas frequência de amostragens usuais.

Tabela 3.3. – Exemplos de frequências de amostragens.

Taxa de amostragem (Hz)	Qualidade do som
192000 Hz	Áudios de DVD, BD-ROM (discos Blu-ray) e HD DVD
176400 Hz	Gravadores HDCD e aplicações profissionais para produção de CDs
44100 Hz	Áudios de CD, MPEG-1 (VCD, SVCD, MP3)
22050 Hz	Áudios PCM e MPEG de baixa qualidade
16000 Hz	Sistemas de comunicação VoIP e VVoIP
8000 Hz	Telefone, walkie-talkies encriptados, (adequado para a fala humana sem sibilância - /s/ soa como /f/)

Fonte: (SAMPLING (SIGNAL PROCESSING), 2017).

3.2.6. Representações de sinais contínuos e discretos

No estudo do processamento de sinais, algumas definições de sinais digitais se fazem necessárias e recorrentes. As mais empregadas seguem definidas abaixo e ilustradas na figura 3.9:

- impulso unitário (ou delta de Kronecker):

$$\delta[n] = \begin{cases} 1, & n = 0 \\ 0, & n \neq 0 \end{cases} \quad (3.41)$$

- degrau unitário:

$$u[n] = \begin{cases} 1, & n \geq 0 \\ 0, & n < 0 \end{cases} \quad (3.42)$$

- sinal retangular:

$$rect_N[n] = \begin{cases} 1, & 0 \leq n < N \\ 0, & n < 0 \text{ ou } n \geq N \end{cases} \quad (3.43)$$

- exponencial real:

$$x[n] = a^n u[n] \quad (3.44)$$

- exponencial complexa:

$$\begin{aligned} x[n] &= a^n u[n] \text{ com } a^n = r^n e^{jn\omega_0} \\ \Rightarrow x[n] &= r^n (\cos n\omega_0 + j \sin n\omega_0) u[n] \end{aligned} \quad (3.45)$$

3.2.7. Sistemas Digitais

Sistema Digital é aquele que recebe um sinal de entrada $x[n]$ e gera uma saída $y[n]$ obedecendo a relação:

$$y[n] = T\{x[n]\} \quad (3.46)$$

Este sistema T será definido como linear se e somente se:

$$T\{a_1 x_1[n] + a_2 x_2[n]\} = a_1 T\{x_1[n]\} + a_2 T\{x_2[n]\} \quad (3.47)$$

assumindo-se quaisquer valores para a_1 e a_2 e quaisquer sinais $x_1[n]$ e $x_2[n]$.

Um sistema será invariante no tempo se obedecer à relação:

$$y[n - n_0] = T\{x[n - n_0]\} \quad (3.48)$$

3.2.8. Operador Convolução e Sistemas Lineares Invariantes no Tempo

É possível verificar que qualquer sinal de tempo discreto pode ser decomposto em impulsos individuais. Na figura 3.10, um sinal $x[n]$ qualquer foi decomposto amostra por amostra, empregando-se a operação:

$$x[k]\delta[n - k] = \begin{cases} x[k], & n = k \\ 0, & n \neq k \end{cases} \quad (3.49)$$

Assim, estabelece-se a *propriedade seletiva do impulso unitário discreto*, em que todo sinal $x[n]$ pode ser definido pela combinação linear de uma sequência arbitrária de impulsos unitários (eq. 3.41) deslocados:

$$x[n] = \sum_{k=-\infty}^{+\infty} x[k]\delta[n - k], \quad k \in \mathbb{Z} \quad (3.50)$$

Supondo-se que seja conhecido o sinal $h_k[n]$ definido como a saída de um sistema cuja entrada é um impulso unitário $\delta[n - k]$, então a resposta $y[n]$ do sistema linear à entrada $x[n]$ será dada por:

$$y[n] = \sum_{k=-\infty}^{+\infty} x[k]h_k[n] \quad (3.51)$$

$$\delta[n - k] \rightarrow h_k[n] \quad (3.52)$$

Se o sinal for invariante no tempo, ou seja, $h_k[n] = h_0[n - k]$, então:

$$y[n] = \sum_{k=-\infty}^{\infty} x[k]h[n - k] = x[n] * h[n] \quad (3.53)$$

onde $*$ representa o operador soma de convolução.

Por definição, o sistema descrito pela equação 3.53 é um sistema Linear Invariante no Tempo (LTI⁴⁷).

O operador integral de convolução é definido por:

$$x(t) = \int_{-\infty}^{\infty} x(\tau)\delta(t - \tau)d\tau \quad (3.54)$$

$$y(t) = \int_{-\infty}^{\infty} x(\tau)h(t - \tau)d\tau = x(t) * h(t) \quad (3.55)$$

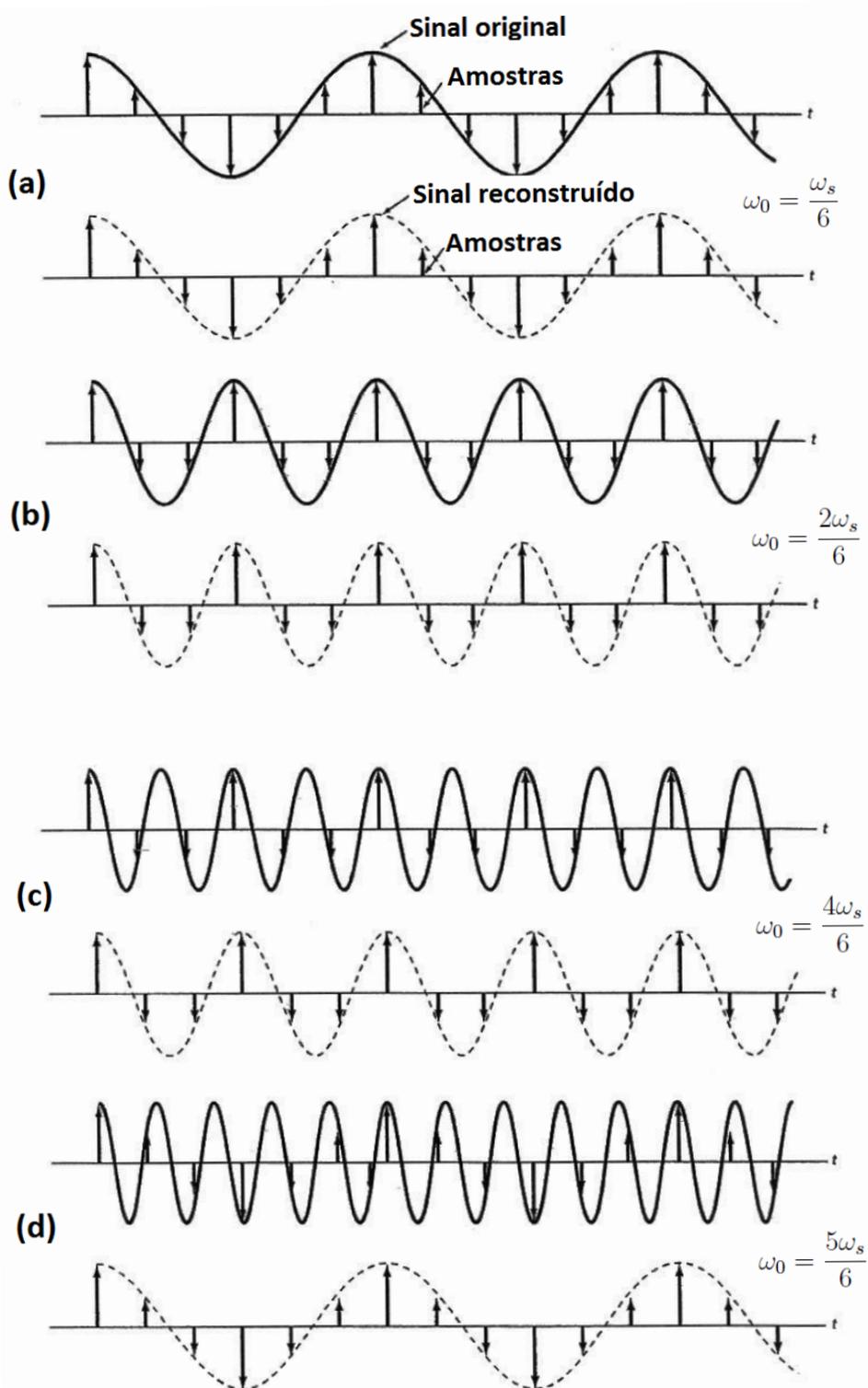
Um sistema LTI obedece às propriedades do operador convolução (tab. 3.4):

Tabela 3.4. – Propriedades do operador soma de convolução.

Comutativa	$x[n] * h[n] = h[n] * x[n]$
Associativa	$x[n] * (h_1[n] * h_2[n]) = (x[n] * h_1[n]) * h_2[n] = x[n] * h_1[n] * h_2[n]$
Distributiva	$x[n] * (h_1[n] + h_2[n]) = x[n] * h_1[n] + x[n] * h_2[n]$

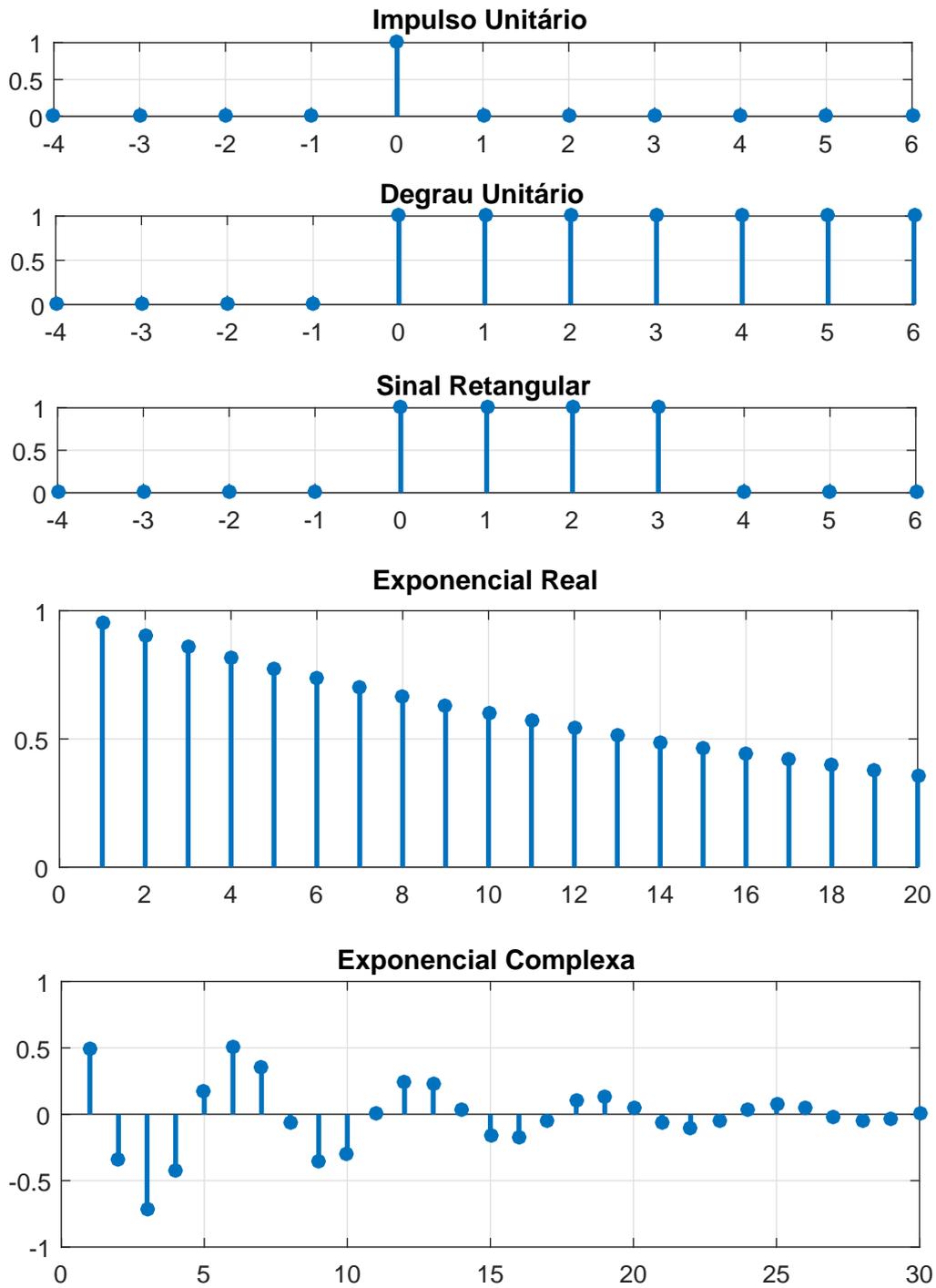
Fonte: (OPPENHEIM; WILLSKY; NAWAB, 2010).

⁴⁷Do inglês, *linear time-invariant*.

Figura 3.8. – Demonstração do efeito de *aliasing*.

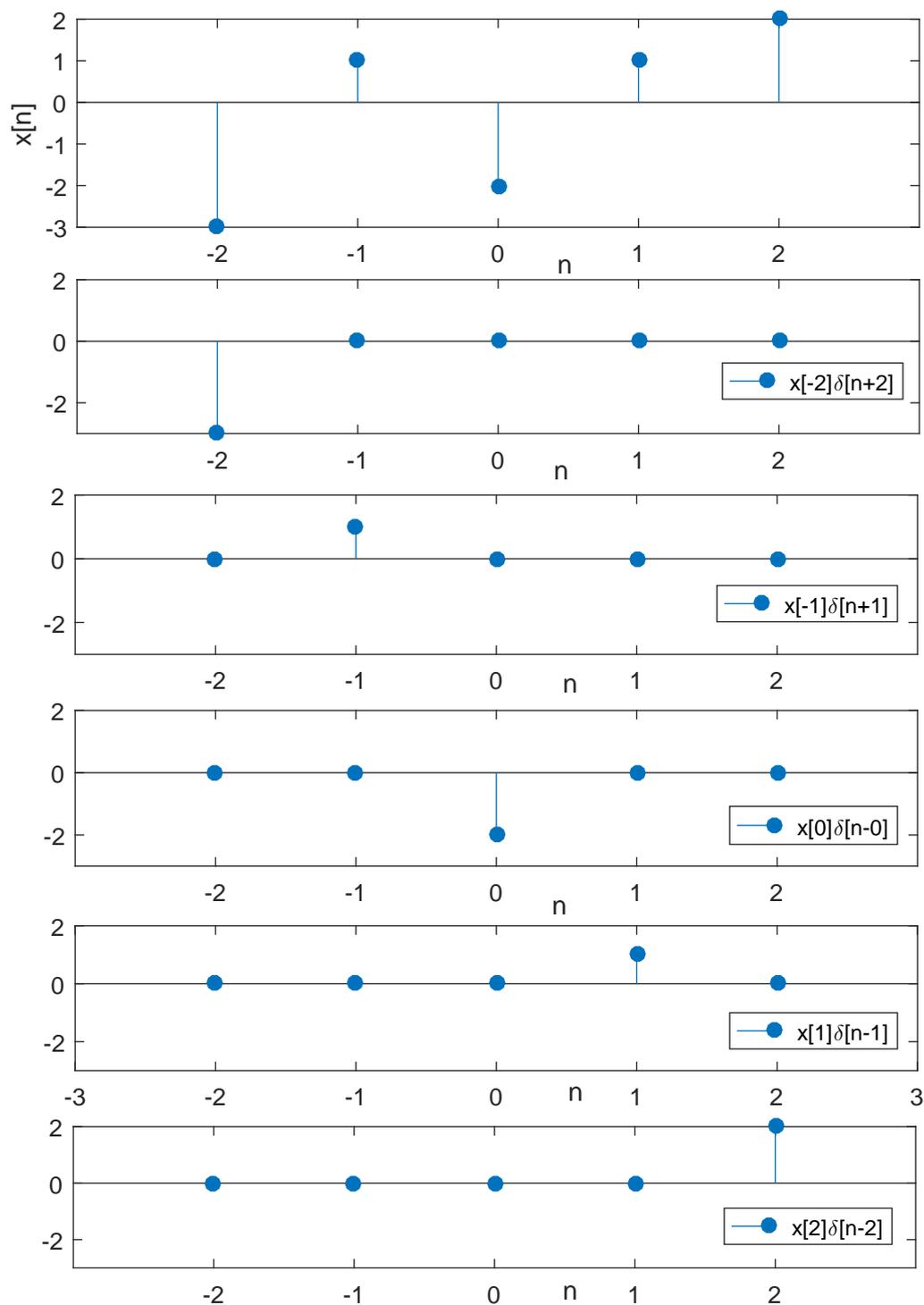
Aliasing em um sinal senoidal. Em cada gráfico, são representados: o sinal original (linha sólida), as amostras (setas), o sinal recuperado via interpolação (linha tracejada) e os valores da frequência de corte ω_0 em função da frequência de amostragem ω_s . Observa-se que em (a) e (b) não ocorre *aliasing*, pois $\omega_0 < \frac{\omega_s}{2}$. Já em (c) e (d), como esta relação não é obedecida, o sinal recuperado não corresponde ao sinal original. Fonte: (OPPENHEIM; WILLSKY; NAWAB, 2010, cap. 7).

Figura 3.9. – Representações de Sinais Discretos.



Representações de sinais digitais discretos conforme equações 3.41 a 3.45. A exponencial real foi produzida com $a = 0,95$ e a exponencial complexa, com $\omega_0 = 1$ e $r = 0,9$. Fonte: produção própria.

Figura 3.10. – Propriedade seletiva do impulso unitário.



No alto, um sinal qualquer $x[n]$. As figuras seguintes representam a decomposição desse sinal em suas componentes individuais nos termos da equação 3.49. Fonte: produção própria.

3.3. Modelos para a voz

Como apresentado na seção 3.1.1, a produção da voz gera uma variação perceptível de pressão no ar, que pode ser registrada através de meios mecânicos e/ou eletrônicos. A voz assim registrada pode ser entendida e analisada como um sinal (sec. 3.2). Na figura 3.4, foi exemplificada uma série temporal de registros de variação de pressão acústica.

Para atender o objetivo de identificar padrões no sinal sonoro que sirvam para identificar falantes, diversos modelos e representações surgiram ao longo do tempo. Dentre os principais estão:

- Modelo Acústico da Produção da Fala: analisa as leis da Física envolvidos na propagação do som no trato vocal. Considera: propagação da onda em três dimensões, formato do trato vocal, perdas por condução de calor e fricção viscosa com as paredes do trato vocal, acoplamento nasal etc.
- Codificação Linear Preditiva (LPC⁴⁸): assume que o sinal da fala é produzido tal qual por uma sirene no final de um tubo (sons vozeados), com ocasionais adições de assobios e sons de estalos (sons sibilantes e explosivos). LPC analisa o sinal da fala pela estimativa dos formantes, removendo seus efeitos do sinal e estimando a intensidade e a frequência da sirene remanescente.
- Representações Perceptualmente Motivadas: conjunto de métodos motivados pelo comportamento do sistema auditório humano

– A Transformada Bilinear

$$s = \frac{z^{-1} - \alpha}{1 - \alpha z^{-1}} \text{ for } 0 < \alpha < 1$$

é um mapeamento no plano complexo que mapea a circunferência unitária nela mesma. É similar às escalas Bark e Mel fazendo-se escolhas apropria-

⁴⁸Do inglês, *Linear Predictive Coding*.

das de α .

- Processamento Cepstral⁴⁹ e Coeficientes Cepstrais de Frequências Mel (MFCC⁵⁰):
- Perceptual Linear Prediction (PLP)
- Rastreamento Estatístico de Formantes: frequências formantes são as ressonâncias no trato vocal. São parâmetros úteis para reconhecimento de falantes, mas bastante difíceis de estimar.

3.3.1. A escala mel

A escala mel⁵¹ foi introduzida em 1937 por Stevens, Volkmann e Newman e é uma escala perceptiva de altura de tom (*pitch*).

Não há uma única equação para a escala Mel e a mais popularizada, a proposta por O’Shaughnessy (2000), é a que está expressa com diferentes bases logarítmicas na equação 3.56 (fig. 3.11).

$$m = 2595 \log_{10} \left(1 + \frac{f}{700} \right) \simeq 1125 \ln \left(1 + \frac{f}{700} \right) \quad (3.56)$$

onde f é a frequência dada em Hz e a expressão de sua inversa é:

$$f = 700 \left[\exp \left(\frac{m}{1125} \right) - 1 \right] \quad (3.57)$$

onde m é a frequência convertida para a escala Mel.

Por ser uma relação empírica, seus resultados são diretamente relacionados ao número de indivíduos entrevistados, à metodologia experimental e à análise dos dados de cada pesquisa. Alguns desses resultados constam das referências (FLETCHER;

⁴⁹Lê-se *cepstral*, pois é um termo formado pela inversão da primeira metade da palavra em inglês *spectral*.

⁵⁰Do inglês, *Mel Frequency Cepstral Coefficients*.

⁵¹O nome vem do inglês *melody* (melodia) para indicar que a escala é baseada em comparações melódicas perceptivas de *pitch*.

Figura 3.11. – Relação entre as escalas Mel e Hertz.

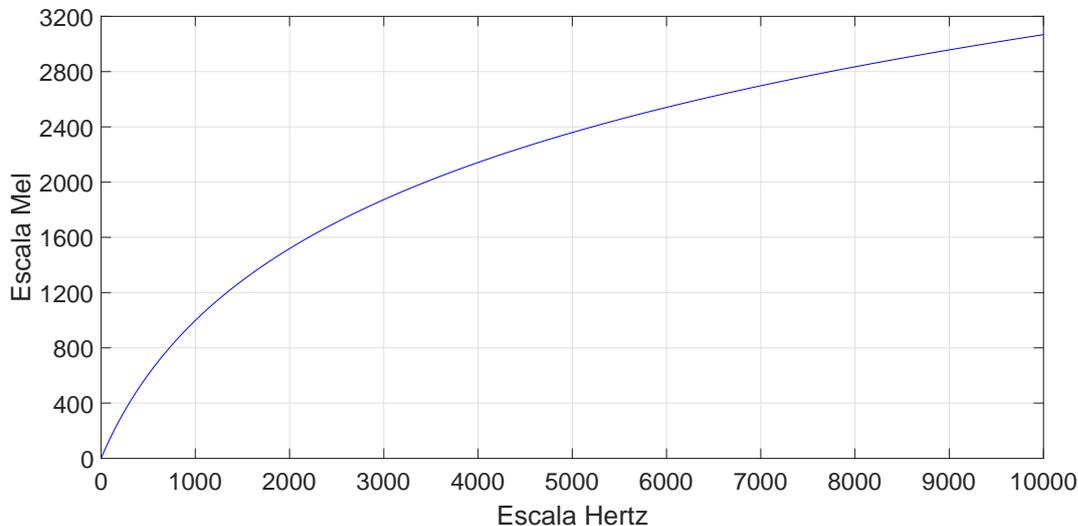


Gráfico da relação entre as escalas Hertz e Mel com base na equação 3.56. Fonte: produção própria.

MUNSON, 1937), (FLETCHER, 1938), (STEVENS; VOLKMANN; NEWMAN, 1937) e (STEVENS; VOLKMANN, 1940).

3.3.1.1. Coeficientes Cepstrais de Frequência Mel (MFCC)

Os Coeficientes Cepstrais de Frequência Mel (MFCC) são parâmetros extraídos de sinais de áudio com a finalidade, dentre outras, de obter características dos falantes com o intuito de identificá-los. São largamente empregados em sistemas automáticos de reconhecimento de fala, pois é bastante robusto contra ruídos de fundo.

Os MFCCs foram introduzidos em 1980 por Davis e Mermelstein e são definidos como o cepstro real de um sinal fragmentado em janelas provenientes do log-espectro de uma FFT (HUANG; ACERO; HON, 2001, p. 314). O procedimento usual para obtenção dos MFCCs segue as etapas:

1. Cortar o sinal em pequenas janelas: tipicamente, janelas de 20 ms a 40 ms com um passo de 10 ms (o que resulta em uma superposição de 10 ms a 30 ms, respectivamente), por exemplo.

2. Calcular a estimativa do periodograma⁵² da potência espectral em cada janela: seja um sinal discreto $x[n]$, seu periodograma é a Transformada de Fourier da estimativa da sequência de autocorrelação. Dado um sinal em que são tomadas amostras à frequência f (amostras por unidade de tempo), o periodograma é definido por:

$$P(f) = \frac{\Delta t}{N} \left| \sum_{n=0}^{N-1} x[n] \exp\{-i2\pi f n\} \right|^2, \quad -\frac{1}{2\Delta t} \leq f \leq \frac{1}{2\Delta t} \quad (3.58)$$

onde Δt é o intervalo de duração temporal da amostra (ou janela). Usualmente, no estudo de voz, emprega-se um filtro do tipo *hanning* – ou do tipo *hamming* – para cada janela.

3. Aplicar o banco de filtros da escala Mel à potência espectral e somar as energias de cada filtro: conforme detalhado à seção 3.3.1.3.
4. Calcular o logaritmo de todas as energias resultantes do banco de filtros
5. Calcular a transformada discreta do cosseno (DCT-II⁵³.) do logaritmo das energias do banco de filtros.
6. Reter os primeiros coeficientes (2 a 13) e descartar o restante.

3.3.1.2. Deltas e delta-deltas

Os coeficientes cepstrais obtidos como acima explicado são também chamados de coeficientes “estáticos” com relação a um segmento ou quadro dentro de uma sequência de quadros, pois representam o estado do sinal em uma determinada janela amostral. A fim de se obter informações sobre a transição entre os parâmetros MFCC, é usual a extração de coeficientes “dinâmicos” denominados **deltas** e **delta-deltas** tomando-se a sequência de vetores MFCC de cada quadro, calculado conforme seção 3.3.1.1.

⁵²Estimativa da densidade espectral.

⁵³Do inglês, *Discrete Cosine Transform* do tipo II.

O cálculo é feito empregando-se uma regressão linear entre os valores de MFCC (GORDILLO, 2013) na seguinte forma:

$$d_t = \frac{\sum_{\theta=1}^{\Theta} \theta (c_{t+\theta} - c_{t-\theta})}{2 \sum_{\theta=1}^{\Theta} \theta^2} \quad (3.59)$$

onde d_t é o coeficiente delta calculado para o instante t , $c_{t-\theta}$ e $c_{t+\theta}$ são os coeficientes MFCC e Θ é o tamanho da janela de regressão.

No presente trabalho, foi adotada a regressão com $\Theta = 2$.

Os parâmetros dinâmicos de segunda ordem denominados delta-deltas (ou duplo-deltas) são obtidos reaplicando-se a eq. (3.59) aos valores de delta anteriormente obtidos.

O vetor empregado para a caracterização do sinal de áudio é construído pelos coeficientes MFCC, seguidos pelos deltas e pelos delta-deltas.

3.3.1.3. Banco de filtros Mel

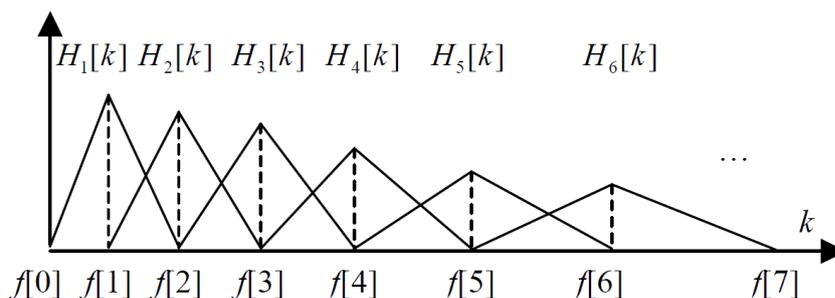
Um banco de filtros é uma superposição de filtros aplicados sequencialmente a um sinal. O banco de filtros Mel é aquele em que a geração de cada filtro procura reproduzir o comportamento da escala Mel incluindo a característica perceptual do ouvido humano de ter melhor discernimento a pequenas variações de *pitch* a frequências baixas e pouca capacidade de distinção na variação de *pitch* em frequências altas.

Um banco de filtros triangulares com M filtros ($m = 1, 2, \dots, M$) é definido por:

$$H_m[k] = \begin{cases} 0 & k < f[m-1] \\ \frac{2(k - f[m-1])}{(f[m+1] - f[m-1])(f[m] - f[m-1])} & f[m-1] \leq k \leq f[m] \\ \frac{2(f[m+1] - k)}{(f[m+1] - f[m-1])(f[m+1] - f[m])} & f[m] \leq k \leq f[m+1] \\ 0 & k > f[m+1] \end{cases} \quad (3.60)$$

e cada filtro retorna a média ponderada espectral em torno de cada frequência central f com largura de banda crescente quanto maior for a frequência. A figura 3.12 apresenta uma aplicação da equação 3.60.

Figura 3.12. – Banco de Filtros Triangulares nos termos da equação 3.60.



Fonte: (HUANG; ACERO; HON, 2001, p. 314).

Uma forma alternativa da equação 3.60 é dada pela equação 3.61:

$$H'_m[k] = \begin{cases} 0 & k < f[m-1] \\ \frac{(k - f[m-1])}{(f[m] - f[m-1])} & f[m-1] \leq k \leq f[m] \\ \frac{(f[m+1] - k)}{(f[m+1] - f[m])} & f[m] \leq k \leq f[m+1] \\ 0 & k > f[m+1] \end{cases} \quad (3.61)$$

que satisfaz $\sum_{m=0}^{M-1} H'_m[k] = 1$. Na prática, o uso de 3.60 ou 3.61 é indiferente, pois essas expressões resultam em cepstros mel divergentes por um vetor constante em todos os valores.

Para a construção do banco de filtros, resta definir a regra de construção dos valores das frequências centrais $f[m]$.

Definindo as frequências inicial e final como $f_i = f[0]$ e $f_f = f[M+1]$, F_S como frequência de amostragem⁵⁴, M o número de filtros e $NFFT$ o tamanho da FFT, as

⁵⁴Comumente, usa-se o subscrito "S", do inglês *sampling*, para amostragem.

frequências $f[m]$, na escala Mel, podem ser obtidas pela equação:

$$f[m] = \left(\frac{NFFT}{F_S} \right) M^{-1} \left(M(f_1) + m \frac{M(f_f) - M(f_i)}{M+1} \right) \quad (3.62)$$

onde M e M^{-1} estão definidos nas equações 3.56 e 3.57, respectivamente.

Outro procedimento, um tanto mais intuitivo para obtenção das frequências $f[m]$ segue abaixo, exemplificado para 10 filtros entre $f_i = 100$ Hz e $f_f = 8000$ Hz :

1. Converter as frequências inicial e final usando a equação 3.56: 100 Hz \simeq 150,22 Mels e 8000 Hz \simeq 2835,00 Mels.
2. Para produzir 10 filtros, são necessárias 12 frequências. Elas devem ser linearmente espaçadas na escala Mel (os valores abaixo estão em Mels):

150,22	394,29	638,36	882,43	1126,50	1370,58
1614,65	1858,72	2102,79	2346,86	2590,93	2835,00

3. Empregando-se a equação 3.57, converte-se os valores acima de volta para Hertz:

100,00	293,83	534,62	833,74	1205,34	1666,98
2240,46	2952,88	3837,92	4937,38	6303,23	8000,00

4. Devido às características de construção da FFT, não há resolução suficiente para calcular os filtros nos exatos pontos da tabela acima, fazendo-se necessário arredondar as frequências para bin da FFT mais próximo. Usa-se a equação:

$$f_{filtros} = \left\lfloor \frac{\left(\frac{NFFT}{2} + 1 \right) f}{F_S} \right\rfloor \quad (3.63)$$

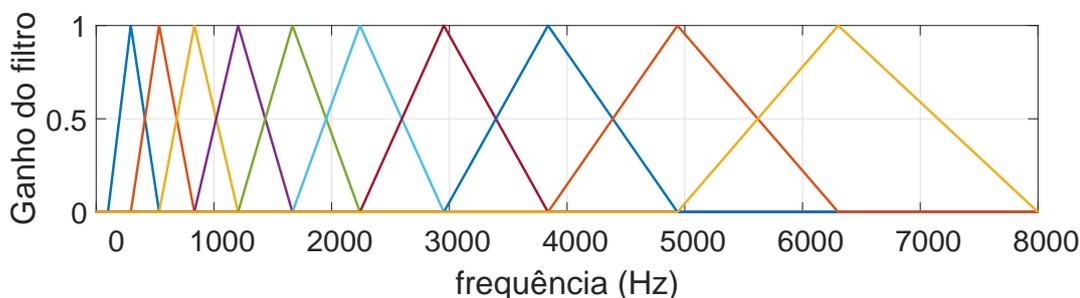
onde f são as frequências obtidas no item 3. No presente exemplo, usando-se $NFFT = 512$, obtêm-se:

2, 8, 16, 26, 38, 52, 70, 94, 122, 158, 202 e 256

5. Efetua-se a adição do log-espectro ponderado pelo banco de filtros, aplicando-se a equação 3.60 ou 3.61 (a última, neste exemplo).

6. O gráfico 3.13 apresenta os filtros obtidos.

Figura 3.13. – Banco de Filtros Mel Triangulares.



Banco de Filtros Mel triangulares nos termos da equação 3.61 com frequências limite 100,00, 293,83, 534,62, 833,74, 1205,34, 1666,98, 2240,46, 2952,88, 3837,92, 4937,38, 6303,23 e 8000,00 (em Hertz).
Fonte: produção própria.

3.4. Análise de Componentes Principais

A Análise de Componentes Principais ou PCA⁵⁵ – como é comumente referido na literatura científica – é um procedimento de regressão linear para redução de variáveis. É muito útil quando foram tomados dados de algumas variáveis (possivelmente muitas) e há suspeita de que haja alguma redundância entre elas. Entende-se essa redundância como havendo correlação entre algumas das variáveis, logo, espera-se que seja possível reduzir o problema ao estudo da composição de apenas algumas variáveis que respondam pela maior parte da variância nos dados.

Outra aplicação é a formação de *clusters* de dados que carregam similaridades entre si, portanto, é ferramenta para separação de dados.

PCA é um método do tipo Análise Fatorial⁵⁶, mas é importante frisar que é conceitualmente distinto deste pois:

- a redução de variáveis por Análise Fatorial restringe-se a rearranjar os dados em função dos fatores – variáveis reais – existentes nos dados, enquanto PCA usa

⁵⁵Do inglês, *Principal Component Analysis*.

⁵⁶Do inglês, *Factor Analysis*.

variáveis artificiais (componentes) que são obtidas com base nas variáveis reais.

- A Análise Fatorial pressupõe a existência de estruturas causais subjacentes: a covariância das variáveis observáveis é devida à presença de uma ou mais variáveis latentes que exercem influência causal nas variáveis observáveis.
- PCA não requer nenhum pressuposto. Trata-se simplesmente de um procedimento para redução de variáveis que (tipicamente) resulta em um relativamente pequeno número de componentes que respondem pela maior parte da variância em um conjunto de variáveis observadas.

3.4.1. Obtenção dos componentes principais

Seja x_1, \dots, x_p o conjunto de variáveis originalmente medidas em um experimento e sejam ξ_i com $i = 1, 2, \dots, p$ combinações lineares quaisquer dessas variáveis:

$$\xi_i = \sum_{j=1}^p a_{ij}x_j \quad (3.64)$$

ou

$$\xi = \mathbf{A}^t \mathbf{x} \quad (3.65)$$

onde ξ e \mathbf{x} são vetores de variáveis aleatórias e \mathbf{A} é a matriz dos coeficientes.

Existem três abordagens básicas para continuar a resolução:

1. Buscar uma transformação ortogonal \mathbf{A} que resulte em novas variáveis ξ que tenham valores estacionários para suas variâncias (método Hotelling),
2. Buscar transformações ortogonais que resultem em variáveis ξ não correlacionadas ou
3. Considerar o problema geometricamente e encontrar a linha pela qual a soma dos quadrados das distâncias perpendiculares seja mínimo, então buscar o plano que melhor se ajuste e assim por diante.

O método Hotelling é explicitado como segue (ver detalhes em (JOLLIFFE, 2002)).

Consideremos a primeira variável $\xi_1 = \sum_{j=1}^p a_{1j}x_j$.

Escolhe-se $\mathbf{a}_1 = (a_{11}, a_{12}, \dots, a_{1p})$ que maximize a variância de ξ_1 , sujeito à restrição $\mathbf{a}_1^t \mathbf{a}_1 = |\mathbf{a}_1|^2 = 1$ (também chamada de 1ª equação normal).

$$\begin{aligned}
 \text{var}(\xi_1) &= E[\xi_1^2] - E[\xi_1]^2 = \\
 &= E[\mathbf{a}_1^t \mathbf{x} \mathbf{x}^t \mathbf{a}_1] - E[\mathbf{a}_1^t \mathbf{x}] E[\mathbf{x}^t \mathbf{a}_1] = \\
 &= \mathbf{a}_1^t (E[\mathbf{x} \mathbf{x}^t] - E[\mathbf{x}] E[\mathbf{x}^t]) \mathbf{a}_1 = \\
 &= \mathbf{a}_1^t \mathbf{V} \mathbf{a}_1
 \end{aligned} \tag{3.66}$$

onde \mathbf{V} é a matriz de covariância de \mathbf{x} .

Encontrar o valor estacionário de $\mathbf{a}_1^t \mathbf{V} \mathbf{a}_1$ restrito à primeira equação normal $|\mathbf{a}_1| = 1$ é equivalente a encontrar o valor estacionário não condicionado de

$$f(\mathbf{a}_1) = \mathbf{a}_1^t \mathbf{V} \mathbf{a}_1 - v(\mathbf{a}_1^t \mathbf{a}_1 - 1) \tag{3.67}$$

onde v é um multiplicador de Lagrange.

Diferenciando com relação a cada componente de \mathbf{a}_1 e igualando a zero, tem-se:

$$\mathbf{V} \mathbf{a}_1 - v \mathbf{a}_1 = 0 \tag{3.68}$$

Para uma solução não-trivial, \mathbf{a}_1 deve ser um autovetor de \mathbf{V} com autovalor v .

A matriz de covariância \mathbf{V} tem p autovalores $\lambda_1, \dots, \lambda_p$, não necessariamente distintos e diferentes de zero, mas eles podem ser ordenados, por exemplo, como $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p \geq 0$.

Como a variância de ξ_1 é

$$\text{var}(\xi_1) = \mathbf{a}_1^t \mathbf{V} \mathbf{a}_1 = v \mathbf{a}_1^t \mathbf{a}_1 = v \quad (3.69)$$

então maximizar a variância é escolher o maior autovalor, aqui definido como λ_1 , e \mathbf{a}_1 é seu autovetor correspondente.

Esse autovetor não será único se o valor de v for uma raiz repetida da equação característica $|\mathbf{V} - v\mathbf{I}| = 0$.

A variável ξ_1 é o primeiro componente principal e tem a maior variância entre quaisquer combinações lineares com as variáveis originais x_1, \dots, x_p .

Para obter o segundo componente principal, $\xi_2 = \mathbf{a}_2^t \mathbf{x}$, escolhe-se coeficientes $a_{2j}, j = 1, \dots, p$, maximiza-se a variância de ξ_2 e impõem-se as restrições:

- o segundo componente principal deve satisfazer à segunda equação normal: $|\mathbf{a}_2| = 1$ e
- ξ_2 não pode ser correlacionado com o primeiro componente principal ξ_1

Essas restrições implicam em

$$E[\xi_2 \xi_1] - E[\xi_2]E[\xi_1] = 0$$

o que leva a

$$\mathbf{a}_2^t \mathbf{V} \mathbf{a}_1 = 0 \quad (3.70)$$

Como \mathbf{a}_1 é autovetor de \mathbf{V} , então garante-se a ortogonalidade entre \mathbf{a}_2 e \mathbf{a}_1 , pois $\mathbf{a}_2^t \mathbf{a}_1 = 0$.

Usando novamente o método dos multiplicadores indeterminados de Lagrange, a maximização da variância passa pela resolução de

$$\mathbf{a}_2^t \mathbf{V} \mathbf{a}_2 - \mu (\mathbf{a}_2^t \mathbf{a}_2 - 1) - \eta \mathbf{a}_2^t \mathbf{a}_1 \quad (3.71)$$

Diferenciando-se com relação a cada termo de \mathbf{a}_2 e igualando-se a zero tem-se

$$\mathbf{V}\mathbf{a}_2 - \mu\mathbf{a}_2 - \eta\mathbf{a}_1 = 0 \quad (3.72)$$

Multiplicando-se por \mathbf{a}_1^t pela esquerda:

$$\mathbf{a}_1^t\mathbf{V}\mathbf{a}_2 - \eta = 0 \quad (3.73)$$

O que mostra que $\eta = 0$, posto que $\mathbf{a}_2^t\mathbf{V}\mathbf{a}_1 = \mathbf{a}_1^t\mathbf{V}\mathbf{a}_2 = 0$.

Logo,

$$\mathbf{V}\mathbf{a}_2 - \mu\mathbf{a}_2 - \eta\mathbf{a}_1 = 0 \quad (3.74)$$

reduz-se a

$$\mathbf{V}\mathbf{a}_2 = \mu\mathbf{a}_2 \quad (3.75)$$

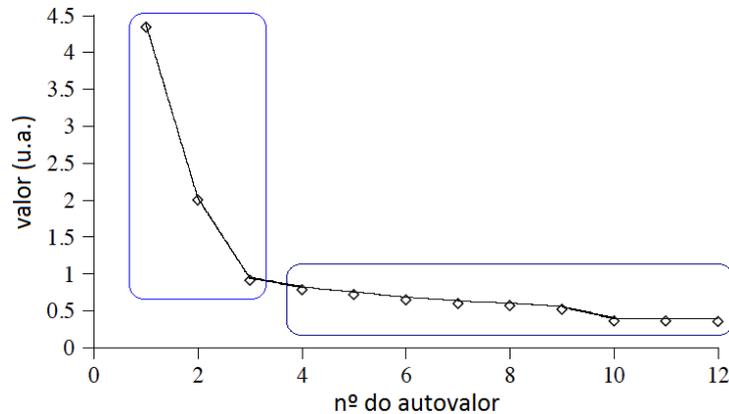
De $\mathbf{V}\mathbf{a}_2 = \mu\mathbf{a}_2$ conclui-se que \mathbf{a}_2 também é um autovetor de \mathbf{V} e é ortogonal a \mathbf{a}_1 . Como o objetivo é maximizar a variância, \mathbf{a}_2 deve corresponder ao segundo maior autovalor, como aqui definido, λ_2 .

O procedimento pode ser repetido até o k-ésimo componente $\xi_k = \mathbf{a}_k^t\mathbf{x}$, onde o autovetor \mathbf{a}_k corresponderá ao k-ésimo maior autovalor de \mathbf{V} com variância igual a este k-ésimo autovalor. Em notação matricial, escreve-se $\xi = \mathbf{A}^t\mathbf{x}$ onde $\mathbf{A} = [\mathbf{a}_1, \dots, \mathbf{a}_p]$ é a matriz cujas colunas são os autovalores de \mathbf{V} .

Em PCA, a qualidade do ajuste pode ser medida com base em alguns critérios. O espectro de autovalores serve para definir quantos componentes principais serão necessários para explicar boa parte da variância dos dados. Quando os 3 primeiros componentes respondem por aproximadamente 80% da variância total, tem-se um caso ideal. Um exemplo típico de espectro de autovalores encontra-se na figura 3.14, em que essa característica é observada e pode-se perceber também que os demais componentes têm praticamente autovalores iguais entre si e de pequena monta se

comparados aos três primeiros componentes.

Figura 3.14. – Exemplo de um “Espectro de Autovalores”.



Por inspeção, percebe-se que os valores dos três primeiros autovalores são bastante superiores aos demais, o que sugere que o ajuste dos dados via PCA é adequado. Fonte: produção própria.

Em resumo, o procedimento explicado define um componente principal como uma combinação linear de variáveis observáveis com pesos otimamente calculados.

3.5. SUVREL: Supervised Variational Relevance Learning

O SUVREL, ou *Supervised Variational Relevance Learning*, é um método variacional de aprendizado por relevância que, através da determinação de um tensor métrico, permite a identificação de padrões baseados em similaridades e foi introduzido por Boareto et al. em 2015. O método apresenta uma função custo que penaliza grandes distâncias entre dados de uma mesma classe e favorece pequenas distâncias entre classes, além de ter a vantagem de permitir sua resolução analítica. O SUVREL é útil como ferramenta de pré-processamento de dados de modo a permitir sua classificação mais eficiente. Trata-se de método supervisionado, pois requer o conhecimento prévio das classes às quais pertence cada dado.

Seja um experimento i e seu correspondente vetor de dados $\vec{x}_i = (x_{i1}, \dots, x_{iF})$

nos espaço dos parâmetros \mathbb{R}^F , e cada parâmetro μ representa uma dimensão com $\mu = 1, \dots, F$. Sejam os dados de aprendizado e validação, \mathcal{L} e \mathcal{V} , organizados em conjuntos de pares $\{\vec{x}_i, c_m\}$ – ou concisamente $\vec{x}_i^{c_m}$ – onde c_m são etiquetas de classe, seja $\mathcal{C} = \{c_1, c_2, \dots, c_m\}$ o conjunto das m classes no experimento, com $m = \{1, 2, \dots, k\}$, e o número total de elementos do conjunto de dados n seja dado por:

$$n = \sum_{m=1}^k n_{c_m} \quad (3.76)$$

onde n_{c_m} é o número de elementos da m -ésima classe⁵⁷.

Uma das maneiras de se definir a geometria do espaço de parâmetros é impor uma estrutura através de um tensor métrico $g_{\mu\nu}$. Seja a distância entre dois elementos i e j definida por:

$$d_{ij}^2 = \sum_{\mu\nu} g_{\mu\nu} \Delta x_{ij\mu} \Delta x_{ij\nu} \quad (3.77)$$

onde $\Delta x_{ij\nu} = x_{i\nu} - x_{j\nu}$ é a diferença dos padrões i e j para o ν -ésimo parâmetro.

Na literatura, há diversos modelos e funções que penalizam os parâmetros que aumentam a distância dos elementos da mesma classe enquanto favorecem parâmetros que diminuem as distâncias entre classes (KULIS, 2013) (BELLET; HABRARD; SEBBAN, 2013). O modelo SUVREL introduz a função custo:

$$E(\{g_{\mu\nu}\}; \mathcal{L}) = \sum_{a \in \mathcal{C}} \langle d_{ij}^2 \rangle_{a,a} - \gamma \sum_{\langle a \neq b \rangle \in \mathcal{C}} \langle d_{ij}^2 \rangle_{a,b} \quad (3.78)$$

onde a e b referem-se a índices de etiquetas de classes para explicitar operações intra e interclasses e γ é um parâmetro que controla o peso dos termos referentes às distâncias interclasses com respeito ao termos das distâncias intraclasses.

A média das distâncias d_{ij}^2 é definida por:

$$\langle d_{ij}^2 \rangle_{a,b} = \frac{1}{n_{ab}} \sum_{i \in a, j \in b} d_{ij}^2 \quad (3.79)$$

⁵⁷Por completeza e para facilitar a compreensão, introduzi o índice m na equação 3.76.

onde $n_{ab} = n_a n_b$ e $\sum_{i \in \mathcal{C}}$ representa uma soma sobre todos os elementos que pertencem à classe a .

A função custo 3.78 pode ser rescrita como:

$$E = \sum_{\mu\nu} g_{\mu\nu} \varepsilon_{\mu\nu} \quad (3.80)$$

onde:

$$\varepsilon_{\mu\nu} = e_{\mu\nu}^{in} - \gamma e_{\mu\nu}^{out} \quad (3.81)$$

com os termos:

$$e_{\mu\nu}^{in} = 2 \sum_{a \in \mathcal{C}} cov(\vec{x}_\mu^a; \vec{x}_\nu^a) \quad (3.82)$$

e:

$$e_{\mu\nu}^{out} = (k-1) \sum_{a \in \mathcal{C}} cov(\vec{x}_\mu^a; \vec{x}_\nu^a) + k^2 cov(\mathbf{m}_\mu; \mathbf{m}_\nu) \quad (3.83)$$

definindo-se $\mathbf{m}_\mu = \{m_{a\mu}\} = \left\{ \frac{1}{n_a} \sum_{i \in a} x_{i\mu} \right\}$ e:

$$cov(\mathbf{m}_\mu; \mathbf{m}_\nu) = \frac{1}{k} \sum_{a \in \mathcal{C}} m_{a\mu} m_{a\nu} - \left(\frac{1}{k} \sum_{a \in \mathcal{C}} m_{a\mu} \right) \left(\frac{1}{k} \sum_{a \in \mathcal{C}} m_{a\nu} \right) \quad (3.84)$$

Do exposto, decorre que:

$$\varepsilon_{\mu\nu} = [2 - (k-1)\gamma] \sum_{a \in \mathcal{C}} cov(\vec{x}_\mu^a; \vec{x}_\nu^a) - \gamma k^2 cov(\mathbf{m}_\mu; \mathbf{m}_\nu) \quad (3.85)$$

que depende apenas dos vetores de dados e da informação sobre a qual experimento cada classe a pertence.

A relevância é obtida empregando-se o método dos multiplicadores de Lagrange para minimizar com relação ao tensor métrico $g_{\mu\nu}$, sujeito à restrição de escala

$\sum_{\mu\nu} g_{\mu\nu}^2 = 1$, para γ fixo. Resolvendo-se:

$$\frac{\delta}{\delta g_{\mu\nu}} \left[E + \theta \left(\sum_{\mu\nu} g_{\mu\nu}^2 - 1 \right) \right] = 0 \quad (3.86)$$

obtem-se:

$$g_{\mu\nu} = \frac{-\varepsilon_{\mu\nu}}{\sum_{\mu'v'} \varepsilon_{\mu'v'}^2} \quad (3.87)$$

A principal diferença na função custo da equação 3.78, introduzida por Boareto et al. (2015), é que ela permite uma resolução analítica – o que reduz consideravelmente o custo computacional quando comparado aos métodos numéricos – de modo que o tensor métrico pode ser obtido explicitamente garantindo-se que este será um tensor definido positivo. Na referência (BOARETO et al., 2015), é provado em detalhes que a condição necessária para que o tensor métrico seja definido positivo é que $\gamma > \gamma^*$ definido por:

$$\gamma^* < \frac{2}{(k-1)} \quad (3.88)$$

Tendo em vista que o SUVREL computa analiticamente o tensor métrico, portanto, empregando menor custo computacional em comparação com métricas obtidas numericamente, parece ser promissor o estudo dos efeitos do SUVREL em conjunção com um classificador como Análise de Componentes Principais (PCA), Multidimensional Scaling (MDS) etc. para comparação de voz, fazendo sua aplicação possivelmente adequada na execução de identificação de falantes em aparelhos de menor capacidade de processamento, por exemplo, celulares.

4. Objetivos Gerais

O assunto “perícia forense de voz” é amplo e compreende diversos temas e pesquisas em aberto. Trata-se de campo de utilidade pública e que sempre é assunto na mídia – basta ver o caso de contestação de autenticidade de gravações realizadas por Joesley Batista em que, segundo ele, o interlocutor é o Presidente Michel Temer (CONTRATADO. . . , 2017) (FRAZÃO, 2017).

Tendo travado interesse por perícia forense de voz, meu objetivo com o presente trabalho é aprofundar e ampliar o conhecimento no âmbito da identificação positiva de falantes, especificamente seguindo o modelo *UBM* (sec. 3.1.3), aplicando metodologia inovadora posto que o *SUVREL* (sec. 3.5) nunca foi utilizado para pré-processamento e tratamento de parâmetros acústicos de voz.

Esquemáticamente, os objetivos podem ser assim sumarizados:

1. Desenvolver e aplicar metodologia para separação de vozes por similaridade e comparação de falantes para fins forenses;
2. Comparar os resultados da etapa 1 com aqueles obtidos utilizando-se o pré-processamento por meio do *SUVREL* – *Supervised Variational Relevance Learning* (seção 3.5);
3. Apresentar os resultados em termos de Razão de Verossimilhanças (seção 2.3.5).

5. Desenvolvimento e experimentos

Sumário

5.1. Análise Comunicativa	141
5.1.1. Razão de distinção	141
5.1.2. Cálculo da razão de distinção empregando resultados do <i>Google</i>	142
5.2. Análise de Voz	144
5.2.1. Experimento preliminar	144
5.2.1.1. Métodos	144
5.2.1.2. Resultados	145
5.2.1.3. Conclusões	145
5.2.2. Experimento intermediário: Supervised Variational Relevance Learning (SUVREL) aplicado à comparação de voz	149
5.2.2.1. Introdução	149
5.2.2.2. Métodos	150
5.2.2.3. Resultados	151
5.2.2.4. Conclusões	159
5.2.3. Experimento final	159
5.2.3.1. Métodos	159
5.2.3.2. Resultados	161
5.2.3.3. Conclusões	167

O projeto inicial deste trabalho, elaborado no primeiro semestre de 2013, previa a divisão das atividades em quatro etapas divididas em três blocos principais:

a) Análise Comunicativa

- Busca de uma identificação vocal por análise estatística da frequência de uso de palavras e/ou letras: nesta abordagem, gravações diversas de diálogos seriam degravados (transcritos) e os arquivos de texto resultantes passariam por análise estatística a fim de revelar a presença de jargões típicos e estruturas de fala que permitissem inferir o grau de instrução do falante e/ou sua inclusão em determinado grupo (como facções criminosas);

b) Análise de Voz

- Determinação de uma base vocal com fundamento no conhecimento *a priori* da fonoaudiologia: pretendia-se seguir a premissa de que a estruturação de análise de voz, fala e linguagem consagrada pela prática de fonoaudiólogos e foneticistas (sec. 3.1.3.1.1) poderia ser transposta de seu caráter subjetivo e qualitativo para algo quantitativo e automatizável;
- Determinação da base vocal pela análise estatística de fator principal: nesta etapa, foi prevista a utilização do método estatístico PCA para separar os “elementos de voz” que seriam obtidos na etapa anterior.

c) Opcional:

- Criação de um programa que automatizasse esse processo.

No decorrer do desenvolvimento das atividades e com o aprofundamento do levantamento bibliográfico, foi identificado que o bloco Análise Comunicativa referia-se, em fato, aos campos da Estilometria e da Linguística e que o projeto inicial aproximava-se bastante do subtema Atribuição de Autoria.

Como será explicado a seguir (sec. 5.1.2), a abordagem comunicativa pretendida não se mostrou frutífera. Já a análise de voz, esta foi bastante alterada e adaptada com relação à ideia apresentada no projeto inicial e, assim, teve resultados bastante promissores, os quais seguem detalhados à seção 5.2 e seguintes.

5.1. Análise Comunicativa

O levantamento bibliográfico sobre Atribuição de Autoria, a saber (BURROWS, 2002), (BINONGO; SMITH, 1999), (STEIN; LIPKA; PRETTENHOFER, 2010), (ZHENG et al., 2006), (ALMEIDA, 2014), (SUN et al., 2012), (RUDMAN, 1997a), (HOLMES, 1994), (GURNEY; GURNEY, 1998), (RUDMAN, 1997b) e (ELLEGÅRD, 1962), revelou que seria necessário adaptar o método da *Razão de Distinção* o qual passo a detalhar.

5.1.1. Razão de distinção

A Razão de Distinção ⁵⁸ foi introduzida em 1962 por Ellegård e foi empregada para estudar a autoria das chamadas “Cartas de Junius” – uma série de panfletos políticos, escritos entre 1769 e 1772, que criticavam o governo do rei George III do Reino Unido – e que consistiam de aproximadamente 150000 palavras. Ellegård organizou uma lista com a frequência de palavras usadas pelo pseudônimo Junius e as classificou positivamente ou negativamente características de Junius comparando-as com a frequência relativa de uso das mesmas palavras por seus contemporâneos. A Razão de Distinção (*RD*) é definida por:

$$RD = \frac{FRJ}{FRNJ} \quad (5.1)$$

onde *FRJ* é a frequência relativa de palavras atribuídas a Junius e *FRNJ* é a frequência relativa de palavras em uma amostra de 1 milhão de palavras de contemporâneos a Junius, mas que não incluísse seus textos.

Embora a metodologia de Ellegård tenha sido criticada (CHRÉTIEN, 1964), principalmente com relação à contagem da frequência que não teria sido rigorosa, mas amostral, a razão de distinção persiste e é usada amplamente para atribuição de autoria.

⁵⁸ *Distinctiveness Ratio*.

5.1.2. Cálculo da razão de distinção empregando resultados do *Google*

Expandindo a ideia de Ellegård – comparação com frequências das palavras usadas por contemporâneos de Junius – aventei a hipótese de que seria possível identificar o grau de instrução de um indivíduo com relação a um “indivíduo médio contemporâneo” representado pela quantidade de resultados de cada palavra no *Google*.

Essa ideia pode ser representada pela seguinte equação (adaptada de eq. 5.1):

$$RDG = \frac{FRI}{FRG} \quad (5.2)$$

onde *RD* é a razão de distinção obtida com informações do *Google*, *FRI* é a frequência de uso de palavras de um indivíduo com relação a um *corpus* de textos de sua reconhecida autoria e *FRG* a frequência relativa de uso de uma palavra pelo *Google*.

A barreira encontrada foi justamente na definição da *FRG*, posto que cada valor de quantidade de resultados apresentado pelo *Google* tem caráter absoluto e não há um *corpus* a servir de referência. Uma tentativa foi construir uma base do tipo “falso texto de mesmas palavras” seguindo o roteiro:

1. Constrói-se uma tabela de *FRI*:

- listam-se as palavras empregadas por um indivíduo em um texto cuja autoria está sendo contestada, anotando suas frequências;
- listam-se as palavras empregadas em um *corpus* de textos de reconhecida autoria do indivíduo em questão, anotando suas frequências.
- calcula-se *DR* para cada palavra

2. Constrói-se uma tabela de *FRG*:

- busca-se no *Google* a quantidade de resultados para cada palavra empre-

gada no texto de autoria contestada

- calcula-se *FRG* com relação ao somatório dos resultados obtidos no *Google* para todas as palavras empregadas no texto de autoria contestada – o que passei a denominar “falso texto de mesmas palavras”

O problema observado é que esse “falso texto de mesmas palavras” adquire caráter enviesado, pois seu tamanho – fictício em sua base – varia em função das palavras que o indivíduo em questão empregou. Alguns testes demonstraram haver extrema variabilidade estatística para casos em que o texto tivesse pequena variação numérica de artigos e/ou preposições, os quais possuem maior quantidade de resultados no *Google*.

Apesar do recurso de pesquisa avançada do *Google* permitir a seleção de idioma e país, ele não diferencia funções gramaticais de palavras e faz com que palavras semanticamente diferentes compartilhem a mesma quantidade de resultados.

Uma outra dificuldade residiu no levantamento dos dados no *Google*. Os inúmeros e consecutivos acessos para a pesquisa faziam com que aquele site identificasse a atividade como um ataque ao servidor DNS⁵⁹ e bloqueasse novos acessos por tempo indeterminado. Para contornar esse bloqueio, o *script crawler* – código para buscas automatizadas – escrito em Matlab precisou ser adaptado para distanciar pesquisas consecutivas com intervalos de tempo sorteados entre 30 s e 2 min. Diante disso, o tempo para o levantamento da base de dados foi aumentado enormemente.

Essa linha de pesquisa foi, em princípio, deixada de lado.

⁵⁹ *Domain Name Server*: servidor de nomes de domínio

5.2. Análise de Voz

5.2.1. Experimento preliminar

A fim de compreender os desafios da abordagem delineada nos Objetivos Gerais (capítulo 4), foi realizado um experimento preliminar.

5.2.1.1. Métodos

Foram delineadas as seguintes etapas:

1. Tomada de dados com a gravação de vozes de 15 voluntários, todos homens, com idades entre 35 e 50 anos. Foi solicitado que lessem uma frase que envolvia a palavra **abracadabra** – a qual tem interesse fonético por possuir 5 fonemas **a** abertos. As gravações foram feitas em um celular Samsung® J7 a uma taxa de aquisição de 44 100 Hz. O ajuste do número de filtros MFCC, foi feito para simular resultados para a taxa de amostragem de 8 000 Hz. Foram realizadas 3 gravações da mesma frase, consecutivamente, de cada voluntário.
2. Fazendo-se uso do programa *Matlab*®, foi desenvolvido um *script* que:
 - separa os trechos de interesse;
 - extrai os parâmetros MFCC (sec. 3.3.1.1) de cada trecho de interesse – foi usado um banco de 40 filtros mel;
 - tabela todos os dados, separa-os estatisticamente empregando PCA (sec. 3.4) e gera gráficos;
 - aplica o SUVREL (3.5) aos dados, reorganizando-os em função dos pesos obtidos com o tensor métrico $g_{\mu\nu}$ (eq. 3.87);
 - separa essa nova tabela de dados estatisticamente empregando PCA e gera gráficos.

3. Análise da variância explicada pelos componentes principais obtidos – que é um indicador de performance desse método.

5.2.1.2. Resultados

As figuras 5.1a a 5.1d apresentam os primeiros resultados obtidos. Trata-se da representação gráfica dos *scores* calculados via PCA – o que significa obter uma nova representação dos dados originais, mas no espaço das componentes principais – para amostras do fonema **a** aberto extraído da sílaba **da** da palavra **abracadabra** (três amostras por indivíduo). No alto, a figura 5.1a mostra a nuvem de *clusters* onde cada cor representa um voluntário. Na figura 5.1b, para efeito de melhor visualização, os mesmos dados são apresentados como esferoides centrados no ponto médio de cada nuvem e com raios calculados em função do desvio padrão dos dados de cada indivíduo em cada um dos três eixos. Já as figuras 5.1c e 5.1d, apresentam os resultados obtidos após o pré-processamento dos dados com o SUVREL.

O PCA gera coordenadas espaciais com ordem de grandeza idêntica ao número de variáveis – neste caso, 40 coordenadas –, mas a representação gráfica, restrita a três dimensões, é suficiente se a soma da variância explicada por essas dimensões for igual ou superior a 80%. A tabela 5.1 relaciona os valores percentuais da variância explicada por cada componente principal.

5.2.1.3. Conclusões

Nota-se o aumento da concentração da variância explicada nos três primeiros componentes, antes e após o SUVREL: de 53,73% para 99,13%. Depreende-se que as componentes principais construídas após o uso do SUVREL permitem elaborar meios para selecionar quais são as frequências que melhor distinguem determinado grupo.

A inspeção visual dos gráficos das figuras 5.1b e 5.1d sugere que a separação dos esferoides é melhor antes do SUVREL. Contraditoriamente, isso não é verdade. Em

5.1b, a informação ali representada refere-se apenas aos três primeiros componentes os quais foram formados pelos MFCCs mas respondem apenas por cerca de 50% da variância total dos dados e podem não representar eficazmente as semelhanças e diferenças entre os falantes. Já em 5.1d, os esferoides em três dimensões respondem por quase 100% da variância total e as similaridades e distinções foram melhor representadas – a gravação do indivíduo M revela sensivelmente que este possui a voz mais aguda dentre os voluntários e, na representação pós-SUVREL, este aparece como um esferoide perceptivelmente isolado.

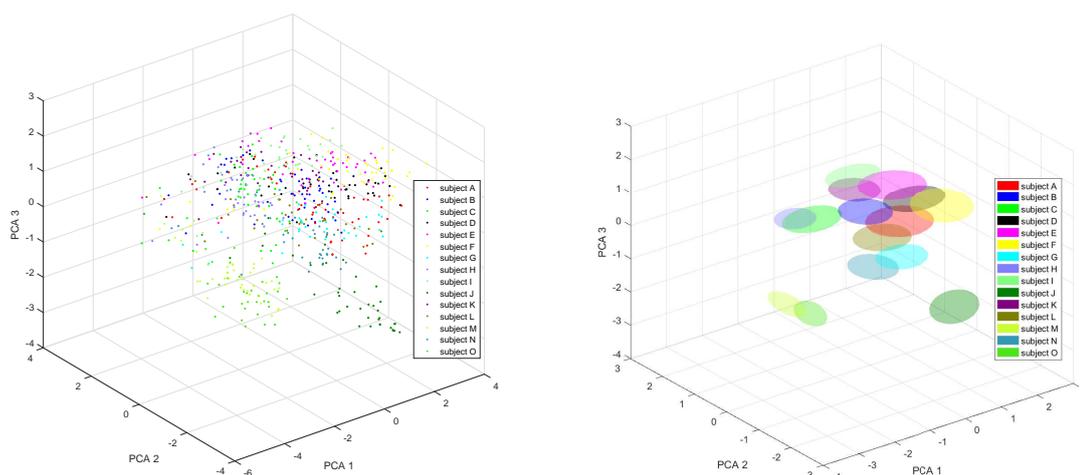
Tabela 5.1. – Variância explicada para dados separados via PCA antes e após pré-processamento com SUVREL.

	Dados brutos		Dados pré-processados com SUVREL	
	variância explicada (%)	somas parciais das variâncias explicadas (%)	variância explicada (%)	somas parciais das variâncias explicadas (%)
PCA1	26,95	26,95	89,45	89,45
PCA2	17,13	44,08	8,74	98,19
PCA3	9,64	53,73	0,94	99,13
PCA4	6,87	60,59	0,39	99,52
PCA5	6,26	66,85	0,16	99,68
PCA6	6,03	72,88	0,14	99,83
PCA7	3,94	76,82	0,09	99,91
PCA8	3,13	79,95	0,03	99,95
PCA9	2,43	82,38	0,02	99,96
PCA10	2,22	84,60	0,01	99,98

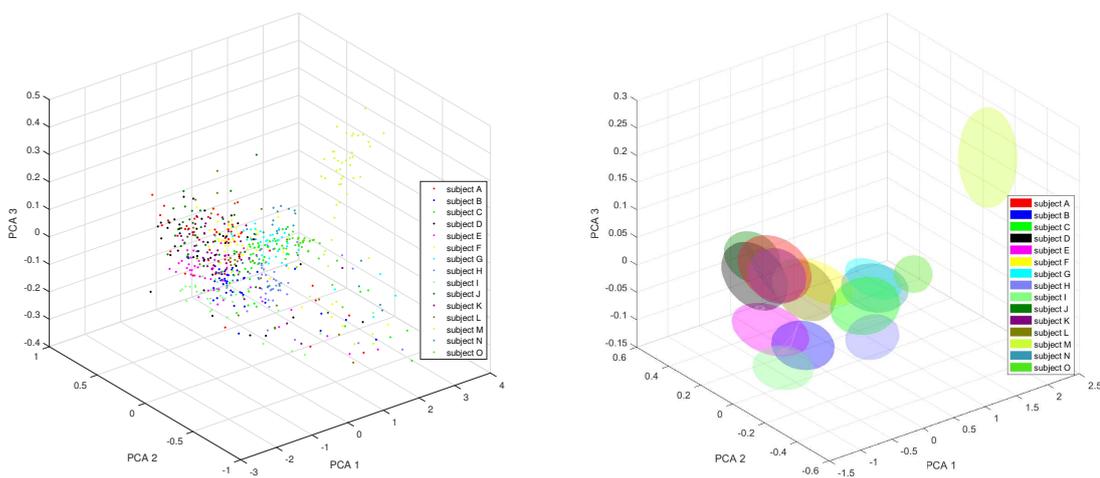
Variância explicada referente aos dados da figura 5.1. Fonte: produção própria.

Procedimento similar ao explicado acima foi feito para a pronúncia completa da palavra **abracadabra**. Os resultados apresentados na tabela 5.2, demonstram significativo aumento da variância explicada, mas a inspeção da figura 5.2 sugere que a qualidade da separação de indivíduos para fins de identificação positiva de falantes é consideravelmente afetada pelo tamanho da amostra. Esse será um dos pontos de investigação a ser aprofundado durante a conclusão do presente trabalho.

Figura 5.1. – Separação de dados via PCA antes e após pré-processamento com SUVREL.



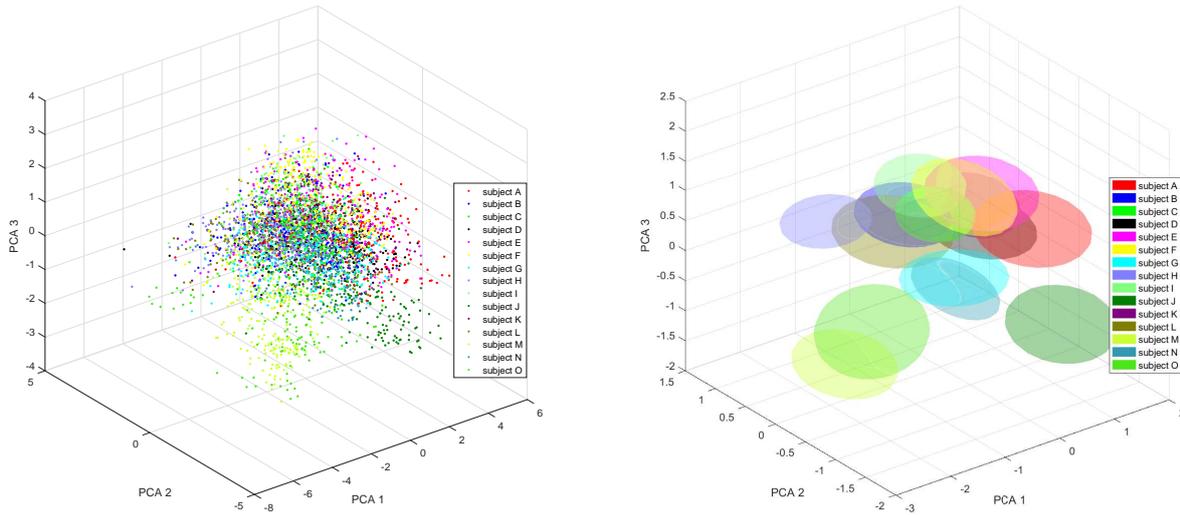
(a) Dados representados no espaço PCA com *clus-terers* sobrepostos (sem SUVREL). Variância explicada: 53,73%.
 (b) Dados no espaço PCA como esferoides representativos do volume médio de abrangência (sem SUVREL). Variância explicada: 53,73%.



(c) Dados representados no espaço PCA com *clus-terers* sobrepostos após pré-processamento com SUVREL. Variância explicada: 99,13%.
 (d) Dados no espaço PCA como esferoides representativos do volume médio de abrangência após pré-processamento com SUVREL. Variância explicada: 99,13%.

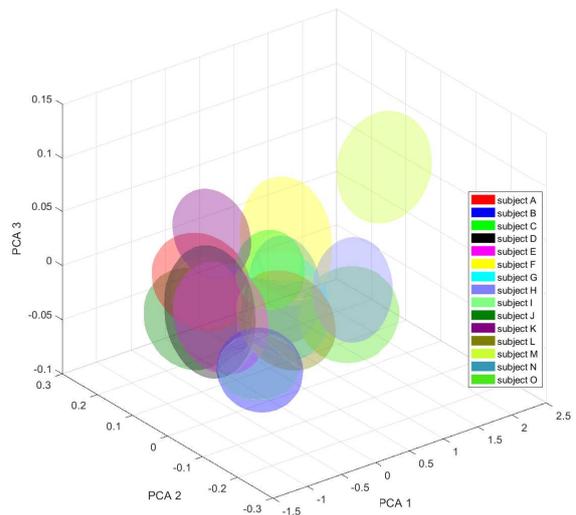
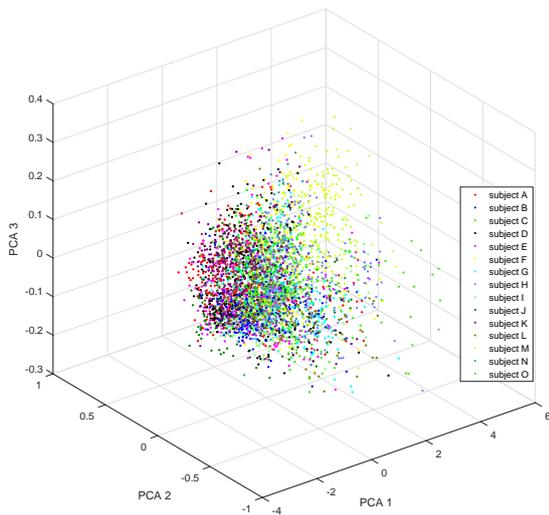
Fonte: produção própria.

Figura 5.2. – Separação de dados via PCA antes e após pré-processamento com SUVREL.



(a) Dados representados no espaço PCA com *clusters* sobrepostos (sem SUVREL). Variância explicada: 48,15%.

(b) Dados no espaço PCA como esferoides representativos do volume médio de abrangência (sem SUVREL). Variância explicada: 48,15%.



(c) Dados representados no espaço PCA com *clusters* sobrepostos após pré-processamento com SUVREL. Variância explicada: 99,25%.

(d) Dados no espaço PCA como esferoides representativos do volume médio de abrangência após pré-processamento com SUVREL. Variância explicada: 99,25%.

Fonte: produção própria.

Tabela 5.2. – Variância explicada para dados separados via PCA antes e após pré-processamento com SUVREL.

	Dados brutos		Dados pré-processados com SUVREL	
	variância explicada (%)	somas parciais das variâncias explicadas (%)	variância explicada (%)	somas parciais das variâncias explicadas (%)
PCA1	24,43	24,43	95,93	95,93
PCA2	14,75	39,18	2,66	98,59
PCA3	8,96	48,15	0,66	99,25
PCA4	7,03	55,18	0,33	99,58
PCA5	5,76	60,94	0,17	99,75
PCA6	5,42	66,35	0,07	99,82
PCA7	4,95	71,30	0,06	99,88
PCA8	3,48	74,78	0,05	99,93
PCA9	2,94	77,71	0,03	99,96
PCA10	2,64	80,35	0,01	99,97

Variância explicada referente aos dados da figura 5.2. Fonte: produção própria.

5.2.2. Experimento intermediário: Supervised Variational Relevance Learning (SUVREL) aplicado à comparação de VOZ

Nesta seção, são mostrados os resultados apresentados na 17^a Conferência Internacional Australasiana sobre Ciência e Tecnologia da Fala (SILVA; TABACNIKS, 2018). Trata-se de evolução do método e dos resultados apresentados à seção 5.2.1.

5.2.2.1. Introdução

A performance de muitos algoritmos para classificação depende fundamentalmente de uma boa métrica no espaço de entrada (*input*). *Support Vector Machines* (SMV), *K-means* e diversas formas de aprendizados de distância supervisionadas do tipo Mahalanobis foram desenvolvidas para atingir este objetivo (BELLET; HABRARD; SEBBAN, 2013). O presente texto consiste em uma pesquisa introdutória quanto à

aplicação do método *Supervised Variational Relevance Learning* (SUVREL) (detalhado à seção 3.5) aplicado à comparação de voz e classificação de um banco de dados com vozes gravadas.

5.2.2.2. Métodos

Esta seção descreve o banco de dados empregado, a extração dos dados e sua análise.

Neste experimento, foi usado um banco de dados de gravações de vozes de mais de 500 falantes de inglês australiano (MORRISON et al., 2015) para fins de pesquisa em comparação forense de voz cujo protocolo de coleta está descrito em detalhes em (MORRISON; ROSE; ZHANG, 2012). Esse banco de dados consiste de três blocos de gravações onde foi solicitado aos falantes que realizassem três tarefas de fala: conversação informal telefônica, troca de informações através do telefone e entrevista em estilo pseudo-policial. Os resultados aqui obtidos foram produzidos com as gravações do primeiro bloco, tanto com indivíduos do sexo masculino, quanto feminino. Cada gravação foi usada em sua frequência original de amostragem de 44 100 Hz. Há resultados comparativos para as gravações em sua alta qualidade original e após a adição de ruído gaussiano branco ($SNR = 25$ dB)⁶⁰. O valor de SNR é dado pela seguinte equação:

$$SNR = 20 \log_{10} \left(\frac{\text{potência do sinal}}{\text{potência do ruído}} \right) \quad (5.3)$$

Usando um *script* desenvolvido em Matlab R2016a®, parâmetros acústicos, especificamente 44 coeficientes cepstrais de frequência Mel (MFCC), foram extraídos para cada indivíduo etiquetado em amostras de 30 segundos de duração divididas em *frames* de 25 milissegundos com superposição de 10 milissegundos, dos quais foram retidos do 2º ao 43º MFCC. O banco de filtros Mel triangular aqui empregado é aquele

⁶⁰Signal to Noise Ratio.

definido em (HUANG; ACERO; HON, 2001, pp. 351).

Os dados formados por cada conjunto de MFCCs, já etiquetados de acordo com cada indivíduo, foram preprocessados com o algoritmo SUVREL. O tensor métrico completo $g_{\mu\nu}$ (eq. 3.87) – em vez de apenas sua diagonal principal – calculado para $\gamma = 2$ fixado (obedecendo à eq. 3.88) foi usado. Tanto os dados originais quanto os preprocessados foram classificados usando Análise de Componentes Principais (PCA) (HAIR JR. et al., 2009). Os percentuais da variância total explicada por cada componente PCA foram então comparados. Gráficos de escores dos três primeiros componentes PCA foram produzidos para apresentar a distribuição dos MFCCs re-escalados no espaço dos componentes principais.

Como o gráfico diretamente obtido com base nos escores apenas levaria a uma nuvem de pontos visualmente incompreensível, e para atingir um melhor entendimento dos resultados, cada “*subnuvem*” de cada indivíduo foi representada por um elipsoide centrado no valor médio da *subnuvem* e esticada proporcionalmente aos desvios padrões, sempre calculados com respeito a cada um dos três componentes principais. Desse modo, cada indivíduo ocupou uma porção de volume no espaço de escores PCA e similaridades entre os indivíduos puderam ser numericamente estimadas pela intersecções nos volumes dos elipsoides. As estimativas na intersecção dos volumes foram calculadas com um algoritmo de Monte Carlo, com taxa de erro inferior a 0.05%. Para ilustrar essas representações, o gráfico apresentado na fig. 5.5 foi feito com apenas um quarto dos valores de desvio padrão a fim de atingir melhor visualização – embora todos os cálculos tenham sido feitos usando-se os valores totais de desvios padrão.

5.2.2.3. Resultados

Os dados de MFCC foram obtidos a partir de amostras de enunciações com duração de 30 seconds (como descrito na seção 5.2.2.2) de cada um dos 100 indivíduos do sexo masculino. Depois da adição de ruído branco gaussiano de $SNR = 25$ dB às

mesmas enunciações, parâmetros MFCC foram novamente obtidos. Esses conjuntos de dados foram pré-processados com SUVREL. Todos os dados foram classificados via Análise de Componentes Principais. A tabela 5.3 apresenta o percentual da variância total explicada pelos primeiros três componentes PCA. Uma melhora considerável depois do pré-processamento com SUVREL e uma concentração de quase toda a variância explicada no primeiro componente PCA podem ser percebidas, o que sugere que a informação de diversos canais de frequência é agora melhor representada. Isto pode ser claramente visto na tabela 5.4, que mostra que a construção do coeficiente do primeiro componente PCA – antes do SUVREL – tem contribuição positiva de alguns poucos MFCCs. Depois do pré-processamento com o SUVREL, uma contribuição positiva pode ser observada de todos os MFCCs o que sugere que este novo espaço de escores não “desperdiça” informações possivelmente úteis ou permite que isto fique espalhado em diversos componentes PCA. Já que mais de 90% de toda a variância explicada está compactada entre os três primeiros componentes, menos esforço computacional para fazer a classificação tende a ser requerido se o SUVREL é adotado como ferramenta de pré-processamento.

Tabela 5.3. – Percentual de variância total explicada pelos primeiros três componentes PCA. Dados de 100 indivíduos do sexo masculino.

	qualidade original	SUVREL	SNR = 25 dB	SUVREL
PCA1	51,71	96,24	60,06	98,17
PCA2	13,79	3,00	12,71	1,70
PCA3	7,68	0,63	6,39	0,08
Total	73,17	99,87	79,15	99,96

Fonte: produção própria.

As figuras 5.3 and 5.4 apresentam gráficos cruzados dos escores PCA entre os eixos dos três primeiros componentes. Enquanto na fig. 5.3, produzida com os dados originais, não é notada nenhuma tendência clara no sentido de alguma organização, na figura fig. 5.4, apresentando os dados pré-processados com SUVREL, é possível

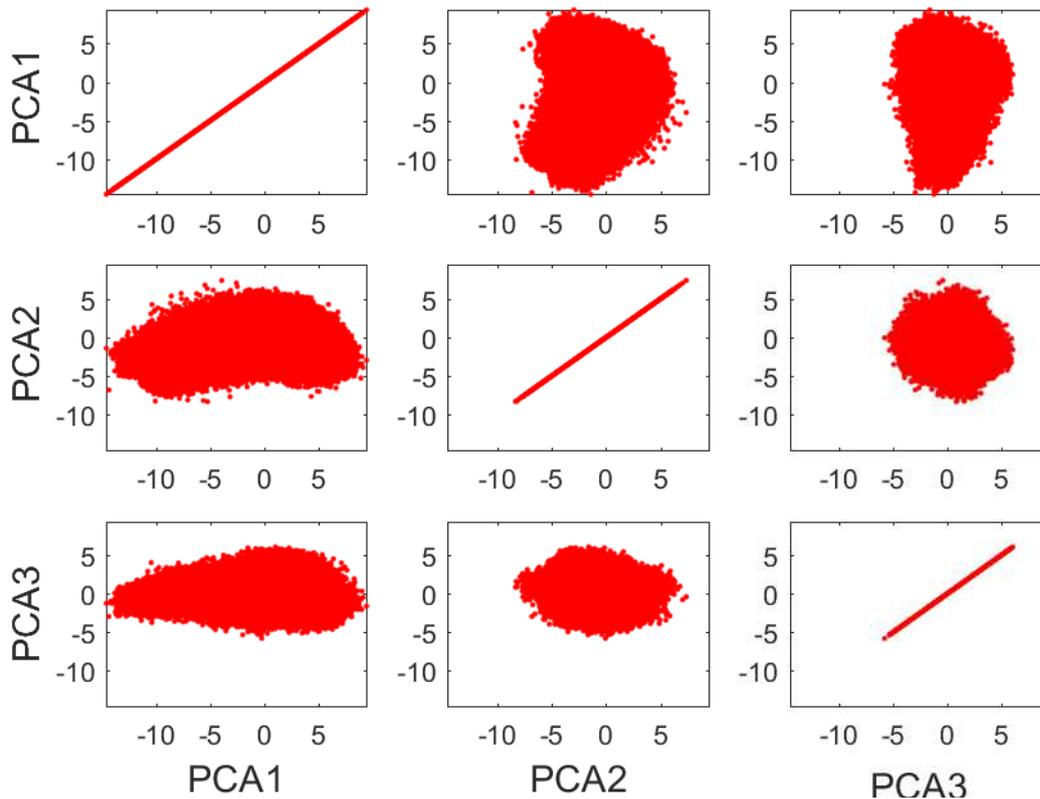
Tabela 5.4. – Contribuição de cada MFCC para o primeiro componente PCA. Dados de 100 indivíduos do sexo masculino.

MFCC	original	SUVREL	MFCC	original	SUVREL
2	0,892	0,512	23	-0,010	0,141
3	0,050	0,184	24	0,000	0,145
4	-0,267	0,042	25	0,002	0,146
5	-0,103	0,107	26	-0,009	0,142
6	-0,127	0,094	27	-0,005	0,144
7	-0,202	0,066	28	-0,002	0,144
8	-0,168	0,078	29	-0,003	0,144
9	-0,071	0,116	30	-0,005	0,144
10	-0,093	0,107	31	-0,003	0,144
11	-0,094	0,107	32	-0,003	0,144
12	-0,048	0,124	33	-0,001	0,145
13	-0,020	0,136	34	-0,005	0,143
14	-0,034	0,131	35	-0,004	0,144
15	-0,033	0,133	36	-0,002	0,145
16	-0,026	0,135	37	0,000	0,146
17	-0,040	0,129	38	0,000	0,145
18	-0,054	0,124	39	0,000	0,146
19	-0,021	0,137	40	0,000	0,146
20	-0,014	0,139	41	-0,001	0,145
21	-0,012	0,141	42	-0,001	0,145
22	-0,014	0,140	43	0,000	0,145

Fonte: produção própria.

notar uma evidente tendência de alinhamento com o eixo do primeiro componente.

Figura 5.3. – Escores PCA.

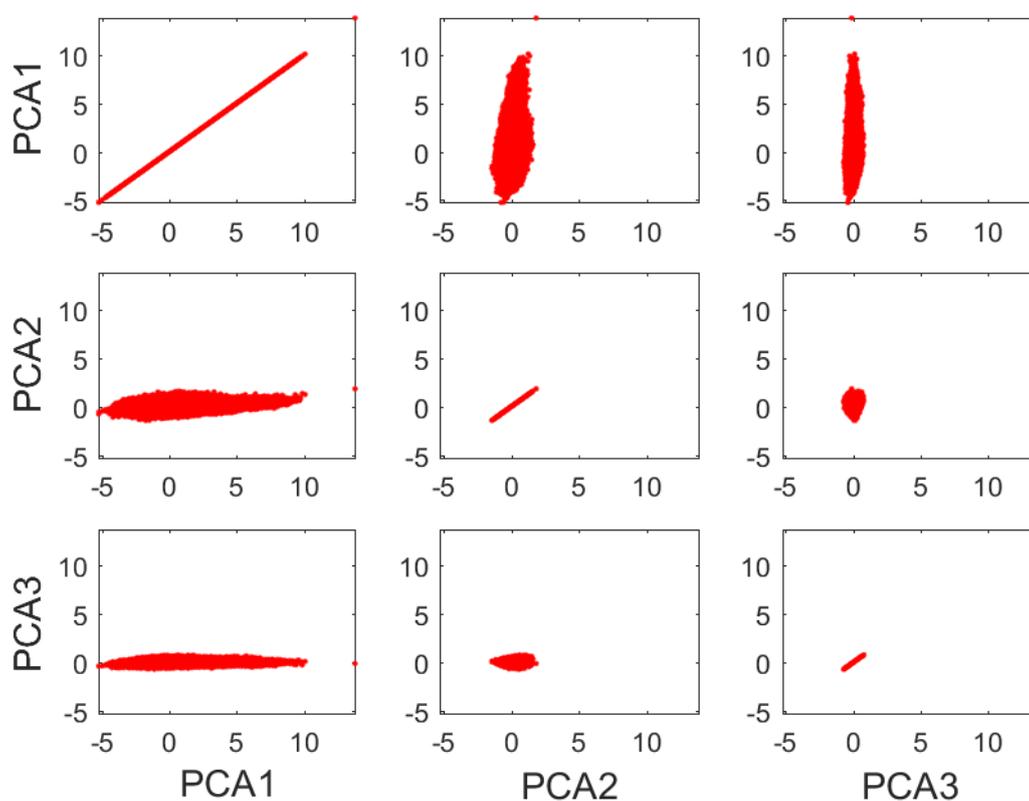


Gráficos cruzados dos escores PCA ao longo dos eixos dos três componentes: nenhuma tendência clara de organização é notada. Fonte: produção própria.

A fim de obter uma melhor visualização da classificação dos indivíduos, fig. 5.5 mostra cada indivíduo como um elipsoide centrado no valor médio do escore PCA e esticado proporcionalmente a um quarto de seu desvio padrão ao longo de cada eixo de componente PCA, após o pré-processamento com SUVREL.

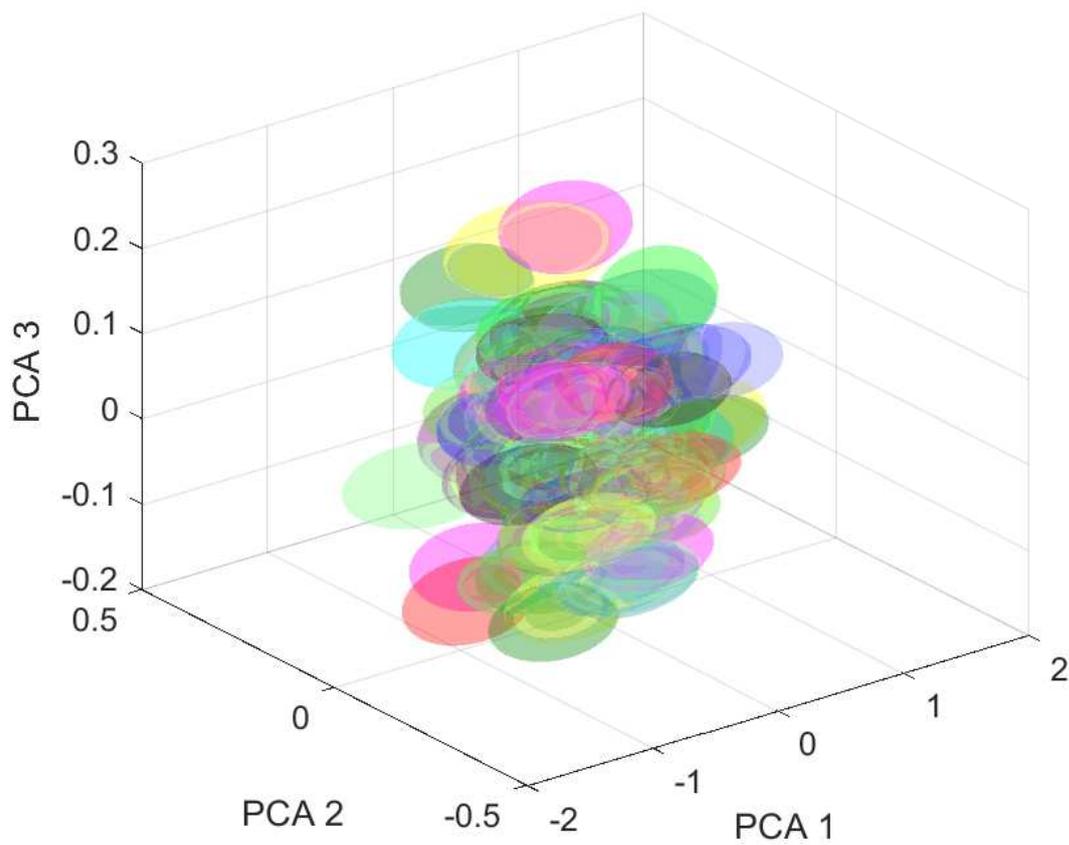
Intersecções volumétricas (usando os valores completos de desvio padrão) foram determinados pelo método de Monte Carlo e cada percentual, com respeito ao próprio volume de cada elipsoide, produziram gráficos: figura 5.6 refere-se ao dado original e a figura 5.7, após SUVREL. Na figura 5.7, foi observado um decréscimo percentual de $20,44 \pm 0,05 \%$ na intersecção volumétrica geral em comparação com a figura 5.6, o que leva a uma melhor classificação de indivíduos.

Figura 5.4. – Escores PCA.



Gráficos cruzados dos escores PCA ao longo dos eixos dos três componentes: mostra uma clara tendência de alinhamento com o eixo do primeiro componente. Fonte: produção própria.

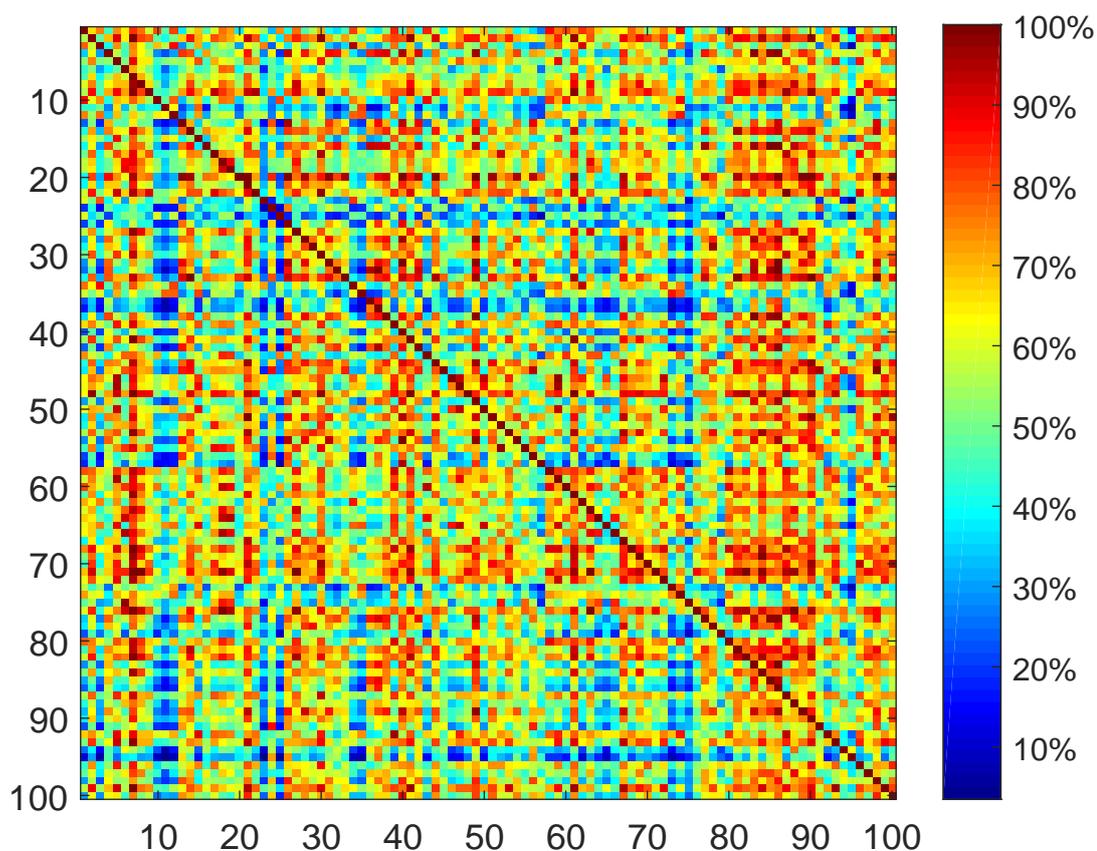
Figura 5.5. – Elipsóides representativos de cada indivíduo.



Elipsóides centrados nos valores médios dos escores PCA e esticados proporcionalmente a um quarto de seus desvios padrão ao longo de cada eixo de componente PCA após pré-processamento com SUVREL. Independentemente das cores repetidas, cada elipsóide representa um indivíduo. Fonte: produção própria.

Análises similares foram realizadas com um conjunto de dados nos quais um ruído branco ($SNR = 25$ dB) foi adicionado. Um decréscimo percentual de $16,22 \pm 0,05$ % na intersecção volumétrica geral foi observada.

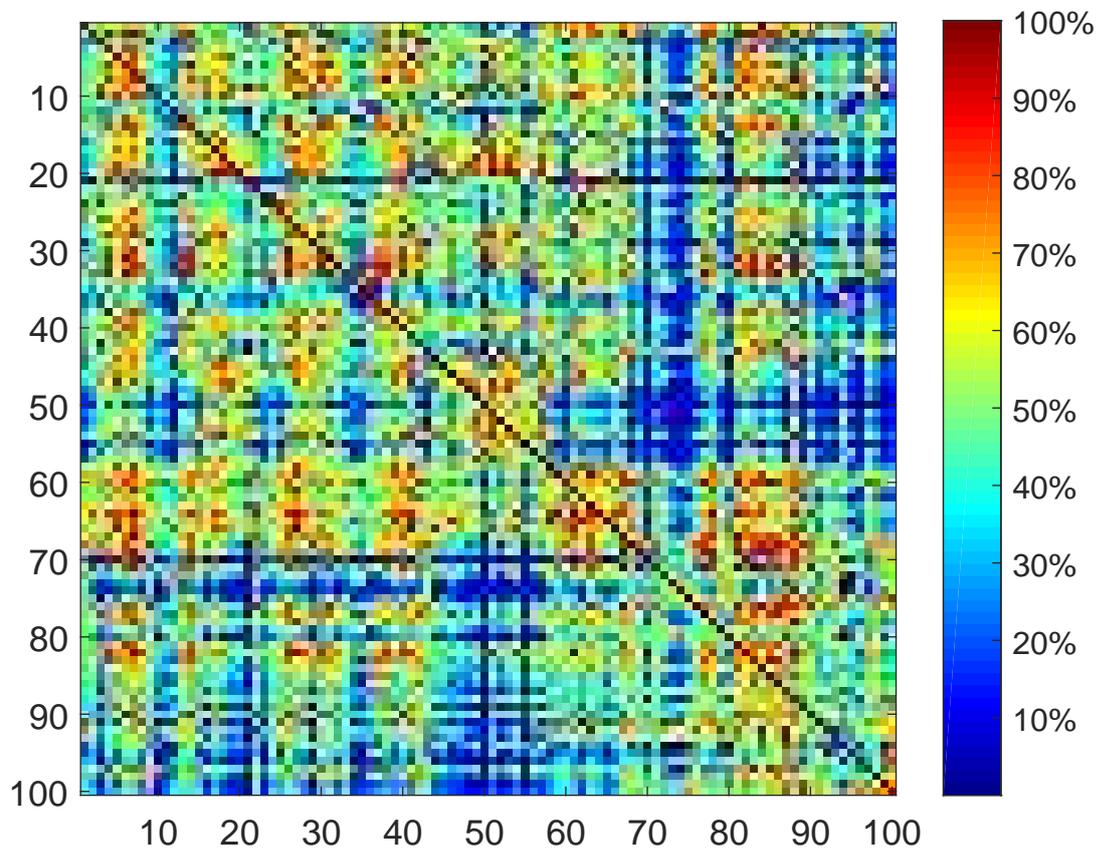
Figura 5.6. – Intersecções volumétricas dos elipsoides representados como temperaturas.



Cada quadrado representa o percentual de volume compartilhado entre elipsoides horizontalmente numerados com relação àqueles verticalmente numerados. Fonte: produção própria.

Também foram obtidos dados de MFCC de amostras de enunciações de 30 seconds (como descrito na seção 5.2.2.2) de cada um dos 100 indivíduos do sexo feminino. A tabela 5.5 mostra o aumento da variância total explicada depois do pré-processamento pelo SUVREL. Na análise dos escores PCA, o percentual dos volumes dos elipsoides resultou em um decréscimo de $6,65 \pm 0,05$ % no caso da qualidade original e de $8,15 \pm 0,05$ % no caso em que o ruído branco foi adicionado.

Figura 5.7. – Intersecções volumétricas dos elipsoides representados como temperaturas.



Cada quadrado representa o percentual de volume compartilhado entre elipsoides horizontalmente numerados com relação àqueles verticalmente numerados. Há um decréscimo percentual de interseção volumétrica geral de $20,44 \pm 0,05 \%$ em comparação com a figura 5.6. Fonte: produção própria.

Tabela 5.5. – Percentual de variância total explicada pelos primeiros três componentes PCA. Dados de 100 indivíduos do sexo feminino.

	qualidade original	SUVREL	SNR = 25 dB	SUVREL
PCA1	48,41	95,68	55,88	98,53
PCA2	14,53	3,54	14,20	1,25
PCA3	8,15	0,63	6,15	0,15
Total	71,09	99,84	76,22	99,93

Fonte: produção própria.

5.2.2.4. Conclusões

O presente trabalho mostra os resultados de uma pesquisa de aplicação prática do método Supervised Variational Relevance Learning (SUVREL) em comparação de voz. Os resultados mostraram que os dados pré-processados com SUVREL levam a uma consideravelmente melhor resolução do classificador, neste caso, a Análise de Componentes Principais.

Como este é um artigo de uma pesquisa introdutória, mais investigações são necessárias para determinar como o pré-processamento via SUVREL afeta a comparação de voz em outras situações, tais como: 1) taxas de frequências de amostragem menores que 44 100 Hz, 2) extração de outros parâmetros, além de MFCC, 3) adição de ruídos mais intensos, 4) diferentes quantidades de indivíduos, 5) comparação com resultados empregando-se apenas a diagonal principal de $g_{\mu\nu}$ (eq. 3.87) etc.

5.2.3. Experimento final

5.2.3.1. Métodos

Neste experimento, foi empregado um banco de dados de gravações de vozes de mais de 500 falantes de inglês australiano (MORRISON et al., 2015) para fins de pesquisa em comparação forense de voz cujo protocolo de coleta está descrito em detalhes em (MORRISON; ROSE; ZHANG, 2012). Esse banco de dados consiste de três blocos de gravações onde foi solicitado aos falantes que realizassem três tarefas de fala: conversação informal telefônica, troca de informações através do telefone e entrevista em estilo pseudo-policial. Os resultados aqui apresentados foram obtidos sobre gravações do primeiro bloco, tanto com indivíduos do sexo masculino, quanto feminino.

A fim de simular as condições de transmissões telefônicas, os arquivos tiveram sua amostragem reduzida de 44 100 Hz para 8 000 Hz – procedimento usualmente

denominado *downsampling*. Foram produzidos resultados com arquivos na qualidade original de gravação do banco de vozes e arquivos que receberam adição de ruído gaussiano branco ($\text{SNR} = 25 \text{ dB}$).

Usando-se um *script* desenvolvido em Matlab R2016a®, foram extraídos para cada indivíduo etiquetado parâmetros acústicos, especificamente 15 coeficientes cepstrais de frequência Mel (MFCC), em amostras de 30 segundos de duração divididas em *frames* de 25 milissegundos com superposição de 10 milissegundos, dos quais foram retidos do 2° ao 15° MFCC. Cada vetor de parâmetros MFCC foi acrescido de seus valores deltas e delta-deltas (seção 3.3.1.2). O banco de filtros Mel triangular aqui empregado é aquele definido em (HUANG; ACERO; HON, 2001, pp. 351).

Os dados formados por cada conjunto de MFCCs, já etiquetados de acordo com cada indivíduo, foram preprocessados com o algoritmo SUVREL. O tensor métrico completo $g_{\mu\nu}$ (eq. 3.87) calculado para $\gamma = 2$ fixado (obdecendo à eq. 3.88) foi usado. Tanto os dados originais quanto os preprocessados foram classificados usando-se Análise de Componentes Principais (PCA) (HAIR JR. et al., 2009).

Os escores dos três primeiros componentes PCA foram empregados para estimar o nível de superposição entre as vozes analisadas. Aplicando-se a abordagem descrita na seção 2.3.5.2, foram obtidos os valores de Razão de Verossimilhanças.

A fim de evitar o enviesamento dos valores calculados de evidência (E), foi padronizado que a gravação denominada “amostra de voz questionada” – conforme esquema das figuras 2.8 a 2.10 (seção 2.3.5.2) – , seria sempre a primeira amostra de voz e a gravação denominada “amostra de voz do suspeito” (banco de dados de controle) seria sempre composta pelas demais amostras de voz do mesmo indivíduo.

Na composição do banco de dados de população relevante, foi usada a seção com maior quantidade disponível de vozes gravadas. Nas tabelas e gráficos a seguir, as quantidades de indivíduos usados como *background* sempre será indicada pela sigla UBM (fazendo referência a *Universal Background Model* como explicado à seção

3.1.3).

Aplicando-se os cálculos delineados à seção 2.3.5.2, foram obtidos os valores de Razão de Verossimilhanças (LR) antes e depois do processamento dos dados com o SUVREL a fim de comparação.

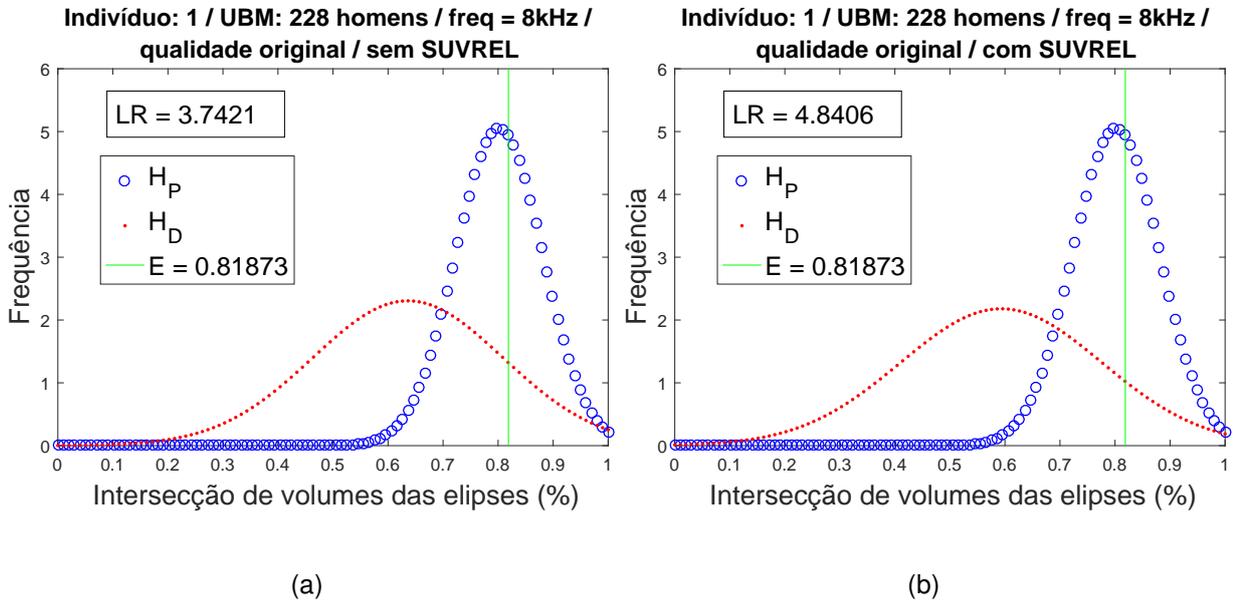
5.2.3.2. Resultados

As tabelas 5.6 a 5.9 relacionam os testes feitos com 20 indivíduos do sexo masculino e 20 indivíduos do sexo feminino, com áudios acrescidos de ruído branco gaussiano ($SNR = 25$ dB) e em sua qualidade original (sem acréscimo de ruído branco). A título exemplificativo, as figuras 5.9 a e 5.9 b trazem os gráficos que produziram os dados para os indivíduos 1 e 2 na tabela 5.6. Todos os gráficos usados para gerar todos os dados das citadas tabelas constam do anexo A.

A interpretação dos resultados passa pela observação dos valores de “variação de LR ” e tendo-se em mente que os testes foram controlados de tal modo que a amostra de voz questionada de cada indivíduo fosse sempre comparada com um banco de dados de controle sabidamente proveniente do mesmo indivíduo, ou seja, resultados “corretos” são aqueles em que a hipótese da promotoria é favorecida, ou seja, resultados que apontem que a amostra de voz questionada e a amostra de voz do suspeito foram geradas produzidas pelo mesmo indivíduo.

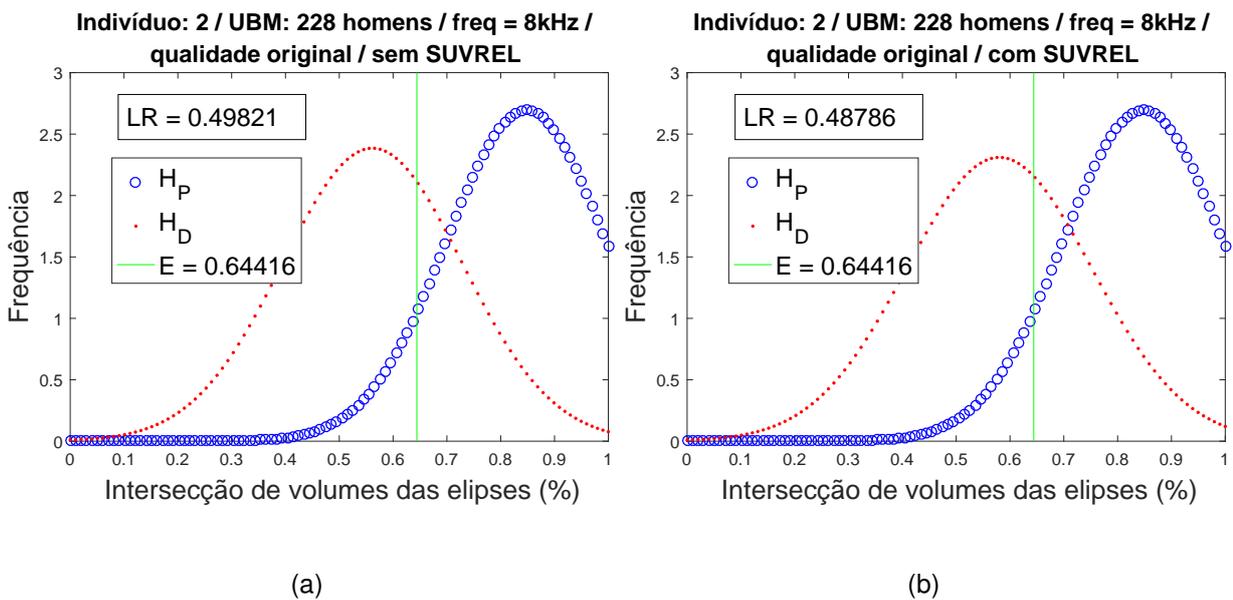
As tabelas foram montadas de modo consistente com a forma fracionária que é proposta para a apresentação de valores de LR , em que valores no numerador representam a força da hipótese da promotoria, enquanto valores no denominador apontam a força da hipótese da defesa (ver seção 2.3.5.3).

Figura 5.8. – Cálculo de LR no cenário apresentado no título de cada gráfico.



FONTE: produção própria.

Figura 5.9. – Cálculo de LR no cenário apresentado no título de cada gráfico.



FONTE: produção própria.

Tabela 5.6. – Cálculos de LR , com e sem SUVREL, com os parâmetros: UBM: 228 homens / freq. = 8kHz / qualidade original.

Indivíduo	Evidência	LR (sem SUVREL)	LR (com SUVREL)	Variação de LR
1	0,82	3,74/1	4,84/1	1,29
2	0,64	1/2,01	1/2,05	0,98
3	0,94	23,64/1	17,71/1	0,75
4	0,77	71,49/1	102,84/1	1,44
5	0,69	1/1,42	1/1,38	1,03
6	0,83	33,83/1	36,35/1	1,07
7	0,70	2,82/1	3,44/1	1,22
8	0,83	4,49/1	4,25/1	0,95
9	0,28	1,19/1	1,08/1	0,90
10	0,80	5,77/1	6,06/1	1,05
11	0,76	1,70/1	2,06/1	1,21
12	0,79	1,23/1	1,46/1	1,19
13	0,87	50,29/1	69,82/1	1,39
14	0,72	1/1,11	1/1,05	1,06
15	0,65	8,27/1	11,24/1	1,36
16	0,67	2,27/1	2,43/1	1,07
17	0,41	1/1,43	1/1,46	0,98
18	0,72	3,23/1	3,38/1	1,05
19	0,53	1,99/1	2,15/1	1,08
20	0,90	1/1,05	1,09/1	1,14

A tabela foi montada de modo consistente com a forma fracionária que é proposta para a apresentação de valores de LR , em que valores no numerador representam a força da hipótese da promotoria, enquanto valores no denominador apontam a força da hipótese da defesa (ver seção 2.3.5.3). O indicador de variação de LR aponta a alteração de tendência dos valores em direção às hipóteses da promotoria (quando a variação é acima de 1) e da defesa (quando a variação é abaixo de 1). Os valores de LR ajustados com o SUVREL não sofreram inversão na força da hipótese apoiada pelos dados, exceto no caso do indivíduo 20, mas trata-se de caso em que o valor de LR gerou resultado inconclusivo ($LR \approx 1$) e a oscilação foi em torno dessa inconclusividade. Todos os gráficos empregados para obter os resultados acima, constam do anexo A.1. Fonte: produção própria.

Tabela 5.7. – Cálculos de LR , com e sem SUVREL, com os parâmetros: UBM: 228 homens / freq. = 8kHz / SNR = 25 dB.

Indivíduo	Evidência	LR (sem SUVREL)	LR (com SUVREL)	Varição de LR
1	0,82	8,14/1	11,18/1	1,37
2	0,66	1/1,34	1/1,28	1,05
3	0,87	77,16/1	63,55/1	0,82
4	0,67	1,52/1	1,92/1	1,26
5	0,58	1,39/1	1,45/1	1,04
6	0,78	13,24/1	16,52/1	1,25
7	0,75	4,70/1	5,82/1	1,24
8	0,79	4,88/1	5,11/1	1,05
9	0,77	1,39/1	1,54/1	1,11
10	0,73	1,20/1	1,20/1	1,00
11	0,50	1,62/1	1,59/1	0,98
12	0,74	1,31/1	1,46/1	1,11
13	0,81	14,79/1	23,79/1	1,61
14	0,74	2,03/1	1,97/1	0,97
15	0,50	2,80/1	3,23/1	1,15
16	0,64	1,71/1	1,65/1	0,97
17	0,28	1/1,40	1/1,54	0,91
18	0,66	1,66/1	1,78/1	1,08
19	0,45	1/1,27	1/1,33	0,95
20	0,74	1,00/1	1,10/1	1,10

A tabela foi montada de modo consistente com a forma fracionária que é proposta para a apresentação de valores de LR , em que valores no numerador representam a força da hipótese da promotoria, enquanto valores no denominador apontam a força da hipótese da defesa (ver seção 2.3.5.3). O indicador de variação de LR aponta a alteração de tendência dos valores em direção às hipóteses da promotoria (quando a variação é acima de 1) e da defesa (quando a variação é abaixo de 1). Os valores de LR ajustados com o SUVREL não sofreram inversão na força da hipótese apoiada pelos dados. Todos os gráficos empregados para obter os resultados acima, constam do anexo A.2. Fonte: produção própria.

Tabela 5.8. – Cálculos de LR , com e sem SUVREL, com os parâmetros: UBM: 319 mulheres / freq. = 8kHz / qualidade original.

Indivíduo	Evidência	LR (sem SUVREL)	LR (com SUVREL)	Variação de LR
1	0,85	10,17/1	12,80/1	1,26
2	0,40	1/3,73	1/4,19	0,89
3	0,75	1,71/1	1,70/1	0,99
4	0,79	2,73/1	3,12/1	1,14
5	1,00	114,71/1	100,34/1	0,87
6	0,78	2,01/1	2,15/1	1,07
7	0,58	2,06/1	2,35/1	1,14
8	0,85	4,02/1	4,33/1	1,08
9	0,67	2,94/1	3,26/1	1,11
10	0,96	1,22/1	1,73/1	1,42
11	0,93	1,89/1	2,06/1	1,09
12	0,85	10,31/1	14,78/1	1,43
13	0,90	17,05/1	11,42/1	0,67
14	0,51	2,03/1	2,66/1	1,31
15	0,70	2,30/1	2,85/1	1,24
16	0,50	1,33/1	1,28/1	0,96
17	0,70	2,22/1	1,98/1	0,89
18	0,80	7,16/1	8,63/1	1,21
19	0,86	29,24/1	11,03/1	0,38
20	0,67	1,69/1	1,74/1	1,03

A tabela foi montada de modo consistente com a forma fracionária que é proposta para a apresentação de valores de LR , em que valores no numerador representam a força da hipótese da promotoria, enquanto valores no denominador apontam a força da hipótese da defesa (ver seção 2.3.5.3). O indicador de variação de LR aponta a alteração de tendência dos valores em direção às hipóteses da promotoria (quando a variação é acima de 1) e da defesa (quando a variação é abaixo de 1). Os valores de LR ajustados com o SUVREL não sofreram inversão na força da hipótese apoiada pelos dados. Todos os gráficos empregados para obter os resultados acima, constam do anexo A.3. Fonte: produção própria.

Tabela 5.9. – Cálculos de LR , com e sem SUVREL, com os parâmetros: UBM: 319 mulheres / freq. = 8kHz / SNR = 25 dB.

Indivíduo	Evidência	LR (sem SUVREL)	LR (com SUVREL)	Varição de LR
1	0,84	9,97/1	11,44/1	1,15
2	0,56	1,30/1	1,46/1	1,12
3	0,65	3,73/1	3,85/1	1,03
4	0,34	1/1,08	1/1,12	0,96
5	0,89	6,17/1	9,34/1	1,51
6	0,73	1,17/1	1,45/1	1,24
7	0,55	1,41/1	1,68/1	1,19
8	0,71	2,04/1	2,69/1	1,32
9	0,71	7,45/1	7,79/1	1,05
10	0,93	1,67/1	2,33/1	1,39
11	0,76	2,89/1	3,03/1	1,05
12	0,77	2,09/1	3,01/1	1,44
13	0,84	9,40/1	8,26/1	0,88
14	0,46	2,37/1	4,51/1	1,91
15	0,56	1,05/1	1,37/1	1,30
16	0,56	1,39/1	1,58/1	1,14
17	0,66	1,49/1	1,42/1	0,96
18	0,70	3,37/1	4,04/1	1,20
19	0,79	5,09/1	5,32/1	1,05
20	0,48	1,03/1	1/1,06	0,92

A tabela foi montada de modo consistente com a forma fracionária que é proposta para a apresentação de valores de LR , em que valores no numerador representam a força da hipótese da promotoria, enquanto valores no denominador apontam a força da hipótese da defesa (ver seção 2.3.5.3). O indicador de variação de LR aponta a alteração de tendência dos valores em direção às hipóteses da promotoria (quando a variação é acima de 1) e da defesa (quando a variação é abaixo de 1). Os valores de LR ajustados com o SUVREL não sofreram inversão na força da hipótese apoiada pelos dados, exceto no caso do indivíduo 20, mas trata-se de caso em que o valor de LR gerou resultado inconclusivo ($LR \approx 1$) e a oscilação foi em torno dessa inconclusividade. Todos os gráficos empregados para obter os resultados acima, constam do anexo A.4. Fonte: produção própria.

5.2.3.3. Conclusões

A inspeção visual das tabelas 5.6 a 5.9 permite concluir que os resultados foram majoritariamente no sentido de apoiar a hipótese da promotoria – de que as amostras de vozes questionada e de fonte conhecida do suspeito são de mesma autoria. Como já explicado, o desenho do experimento definiu esse resultado como sendo o esperado.

A aplicação do SUVREL resultou em ajuste da estimativa de LR e majoritariamente não houve inversão na força da hipótese apoiada pelos dados, à exceção do indivíduo 20 da tabela 5.6 e do indivíduo 20 da tabela 5.9. Entretanto, tratam-se de casos em que os valores de LR geraram resultados inconclusivos ($LR \approx 1$) e a oscilação ocorreu justamente em torno dessa inconclusividade. Não foi possível identificar uma tendência ou uma correlação entre os valores de evidência e a variação de LR , o que permite inferir que os dados pré-processados pelo SUVREL sofreram um ajuste não tendencioso. O SUVREL tem por característica fundamental ser uma função custo que penaliza grandes distâncias entre dados de uma mesma classe e favorece pequenas distâncias entre classes.

Não foi identificada qualquer tendência para que valores de evidência baixos (menores que 0,5) – possível indicativo de maior intravariabilidade na produção de voz daquele indivíduo – sejam associados a aumentos expressivos de valores de LR pós-SUVREL.

A modelagem do experimento empregou apenas técnicas automatizadas de separação e classificação de vozes. Assim, ausente uma etapa de verificação humana por meio de ouvintes treinados em *auditory skills* (seção 3.1.3), a interpretação honesta dos dados das tabelas 5.6 a 5.9 permite apenas concluir que o pré-processamento via SUVREL resultou em um ajuste nas estimativas de LR e que, majoritariamente, esse ajuste se deu no sentido de apoiar a hipótese da promotoria, como era esperado diante do desenho do experimento. Não foi identificado o motivo para que, em alguns casos, houvesse variação das estimativas de LR em sentido oposto – diminuindo a força da

hipótese da promotoria. De todo modo, não se tratam de casos inesperados dentro do que se propõe na metodologia aqui empregada.

6. Conclusões

Propus-me, inicialmente, a realizar pesquisa no campo da identificação positiva automática de falantes por meio de duas abordagens distintas: análise comunicativa e análise de voz. Neste capítulo, retomarei as conclusões já apresentadas na seção “5 Desenvolvimento e experimentos”.

Durante o desenvolvimento da pesquisa em análise de voz, deparei-me com a contemporânea quebra de paradigma, em especial em países da Europa Ocidental, para que peritos forenses prefiram a apresentação de resultados periciais avaliativos (grupo de perícias que requerem uma avaliação por parte do perito a respeito do peso – ou força – da evidência) usando a Razão de Verossimilhanças (que decorre do Teorema de Bayes). Ao aprofundar a pesquisa, notei que discussão semelhante praticamente inexistente no cenário da perícia técnico-científica brasileira. Além disso, não encontrei qualquer artigo escrito que discutisse a aplicabilidade do Teorema de Bayes no contexto de uma legislação em que seja consolidado o *princípio da não culpabilidade*. Neste capítulo, retomarei as conclusões inéditas delineadas no capítulo “2 Aspectos gerais da perícia criminal no ordenamento jurídico brasileiro e aplicação do Teorema de Bayes para a apresentação de seus resultados”.

Serão também citadas as contribuições sociais resultantes da pesquisa aqui produzida e as sugestões de aprofundamentos futuros.

6.1. Abordagem por análise comunicativa

O desenvolvimento das atividades para análise comunicativa foi formulado com base na razão de distinção aplicada em dados de busca no *Google* (equação (5.2)). Essa abordagem mostrou-se infrutífera. O problema observado foi que o “falso texto de mesmas palavras” (um dos parâmetros necessários na equação (5.2)) adquire caráter enviesado, pois seu tamanho – fictício em sua base – varia em função das palavras que o indivíduo em questão empregou. Alguns testes demonstraram haver extrema variabilidade estatística para casos em que o texto tivesse pequena variação numérica de artigos e/ou preposições, os quais possuem maior quantidade de resultados no *Google*. Apesar do recurso de pesquisa avançada do *Google* permitir a seleção de idioma e país, ele não diferencia funções gramaticais de palavras e faz com que palavras semanticamente diferentes compartilhem a mesma quantidade de resultados. Outra dificuldade residiu no levantamento dos dados no *Google* já que acessos consecutivos para a pesquisa faziam com que o site identificasse a atividade como um ataque ao servidor DNS e bloqueasse novos acessos por tempo indeterminado. Para contornar esse bloqueio, o *script crawler* escrito em Matlab precisou ser adaptado para distanciar pesquisas consecutivas com intervalos de tempo sorteados entre 30 s e 2 min. Diante disso, o tempo para o levantamento da base de dados foi aumentado enormemente.

6.2. Abordagem por análise de voz

Foram realizados três experimentos.

O primeiro experimento permitiu identificar a viabilidade da aplicação do SU-VREL em conjunto com o classificador Análise de Componentes Principais (PCA). Os resultados demonstraram uma maior concentração da variância total explicada no espaço dos escores PCA quando foram comparados os dados antes e após a etapa de pré-processamento pelo SU-VREL.

No experimento seguinte, foi posta à prova o achado do experimento anterior quando aplicado a um volume maior de dados. A aplicação a 100 vozes masculinas e a 100 vozes femininas demonstrou a viabilidade do método. Os resultados mostraram que os dados pré-processados com SUVREL levaram a uma consideravelmente melhor resolução do classificador PCA.

No último experimento, a abordagem procurou simular as condições de casos reais de comparação de voz, definindo-se elementos como “voz do suspeito”, “amostra de voz questionada”, “bando de dados de controle”, “banco de dados de população relevante” e “evidência”. Além disso, os resultados foram apresentados na forma de Razão de Verossimilhanças.

Os resultados foram majoritariamente no sentido de apoiar a hipótese da promotória – de que as amostras de vozes questionada e de fonte conhecida do suspeito são de mesma autoria –, conforme era esperado tendo-se em vista o desenho do experimento.

A aplicação do SUVREL resultou em ajuste da estimativa de LR e majoritariamente não houve inversão na força da hipótese apoiada pelos dados. Não foi possível identificar uma tendência ou uma correlação entre os valores de evidência e a variação de LR , o que permite inferir que os dados pré-processados pelo SUVREL sofreram um ajuste não tendencioso. Também não foi identificada qualquer tendência para que valores de evidência baixos (menores que 0,5) – possível indicativo de maior intra-variabilidade na produção de voz daquele indivíduo – sejam associados a aumentos expressivos de valores de LR pós-SUVREL. A modelagem do experimento empregou apenas técnicas automatizadas de separação e classificação de vozes. Assim, ausente uma etapa de verificação humana por meio de ouvintes treinados em *auditory skills* (seção 3.1.3), a interpretação honesta dos resultados permite apenas concluir que o pré-processamento via SUVREL resultou em um ajuste nas estimativas de LR e que, majoritariamente, esse ajuste se deu no sentido de apoiar a hipótese da promotória, como era esperado diante do desenho do experimento. Não foi identificado o motivo

para que, em alguns casos, houvesse variação das estimativas de LR em sentido oposto – diminuindo a força da hipótese da promotoria. De todo modo, não se tratam de casos inesperados dentro do que se propõe na metodologia aqui empregada.

6.2.1. Aprofundamentos futuros

Não foi identificado o motivo para que, em alguns casos, houvesse variação das estimativas de LR em sentido oposto ao que a modelagem do experimento impunha – diminuindo, em vez de reforçar, a força da hipótese da promotoria. Sugere-se pesquisa futura associando os resultados obtidos àqueles conseguidos por meio de ouvintes treinados em *auditory skills* (seção 3.1.3).

Além disso, sugere-se que seja investigado o comportamento do SUVREL quando associado a outros classificadores, bem como, sejam variados os parâmetros de voz para outros além do MFCC.

6.3. Teorema de Bayes na apresentação de Resultados Periciais e sua relação com o Ordenamento Jurídico Brasileiro

Como já citado, durante o desenvolvimento da pesquisa em análise de voz, deparei-me com a contemporânea quebra de paradigma para que peritos forenses prefiram a apresentação de resultados periciais avaliativos (grupo de perícias que requerem uma avaliação por parte do perito a respeito do peso – ou força – da evidência) usando a Razão de Verossimilhanças (que decorre do Teorema de Bayes) em detrimento a escalas de comparação subjetivas que são mais tradicionalmente utilizadas.

Ao aprofundar a pesquisa, notei que discussão semelhante praticamente inexistia no cenário da perícia técnico-científica brasileira. Nada foi encontrado na pesquisa

bibliográfica em língua portuguesa quanto à aplicabilidade e/ou aceitabilidade do Teorema de Bayes no contexto legal brasileiro.

O aprofundamento da pesquisa revelou que em países signatários da Rede Europeia de Institutos de Ciência Forense (European Network of Forensic Science Institutes – ENFSI) (seção “2.1.4 Aceitabilidade Internacional do Teorema de Bayes na apresentação de resultados periciais”) já existe recomendação para que os resultados periciais avaliativos sejam preferencialmente apresentados em termos de Razão de Verossimilhanças.

Além disso, chamou-me a atenção de que já haja ampla argumentação em artigos e livros em língua inglesa rebatendo as críticas legais, que já foram levantadas há anos nos sistemas judiciais onde vigora o *princípio da presunção de inocência*, comumente descrito pelo axioma “inocente até que se prove culpado”. O que se questiona é uma possível incongruência entre a aplicação simultânea desse princípio legal e do arcabouço teórico do Teorema de Bayes. Argumenta-se que a *presunção de inocência* requer do julgador a aceitação inicial tácita da inocência completa do réu e isso vai de encontro a um fundamento básico da aplicação do Teorema de Bayes de que não é possível calculá-lo se a hipótese inicial (H) – *a priori* – for nula ($P(H) = 0$) ou total ($P(H) = 1$): é **impossível** revisar uma posição que inicialmente consista na certeza absoluta sobre algo.

Acontece que o ordenamento jurídico brasileiro consagrou o *princípio da presunção da não culpabilidade*, o qual é sutilmente diferente do *princípio da presunção de inocência* (detalhes na seção “2.1.3 Presunção da não culpabilidade ou presunção de inocência?”).

De modo sucinto, a Constituição Federal afirma que “ninguém será considerado **culpado**(...)” – sem haver presunção de inocência – até o trânsito em julgado da sentença condenatória.

A diferença primordial será no momento do marco final: o marco final dessa

presunção de inocência nos tratados internacionais dos quais o Brasil é signatário se dá com a sentença condenatória em segunda instância, quando há a confirmação – por parte de um órgão colegiado – daquela condenação imposta a uma pessoa. Já no princípio da da presunção não culpabilidade – adotado na Constituição Brasileira –, o marco final é indubitavelmente o trânsito em julgado da sentença condenatória. Assim, no Brasil, alguém só será considerado culpado – não necessariamente será considerado inocente durante esse intercurso – a partir da sentença condenatória transitada em julgado, ou seja, quando a sentença condenatória não mais puder ser recorrida segundo as normas processuais vigentes.

Lançando-se mão da linguística – especificamente da Metalinguagem Semântica Natural –, Rose (2002, p. 72) discute as diferenças entre os componentes semânticos dos verbos *presumir* e *acreditar* e conclui que a presunção de inocência não é uma crítica válida ao uso legal do teorema de Bayes.

Dentro dos limites da pesquisa bibliográfica realizada, realizei abordagem inovadora ao fazer a extensão da discussão acima para abarcar o preceito da *presunção da não culpabilidade*.

Concluí que a argumentação acima pode ser diretamente estendida ao preceito legal da *presunção da não culpabilidade*, pois seu cerne recai sobre o mesmo verbo epistêmico (*presumir*) e apenas a presença da dúvida mínima *a priori* contida no componente semântico *Eu não posso dizer: “Eu sei que Y não é culpado”* já é condição suficiente para a validade da aplicação jurídica do Teorema de Bayes no contexto do ordenamento jurídico brasileiro.

6.4. Contribuições científicas inovadoras

As contribuições científicas inovadoras contidas nesta tese são:

1. Discussão a respeito da aplicabilidade da Razão de Verossimilhanças no ordena-

mento jurídico brasileiro e sua compatibilidade com o *princípio da não culpabilidade* e

2. Investigação sobre o impacto da aplicação da ferramenta de física matemática SUVREL (Supervised Variational Relevance Learning) na separação automática de vozes associada ao classificador Análise de Componentes Principais (PCA).

6.5. Impacto social

A pesquisa de conteúdo a respeito do Teorema de Bayes e da Razão de Verossimilhanças mostrou-se, em grande parte, com linguagem e abordagens inacessíveis ao público geral, seja por tratarem o assunto com foco no formalismo matemático, que dificulta a compreensão prática da técnica, seja pelo excesso de simplificações do uso técnico, que deixam a desejar no entendimento mínimo do formalismo matemático necessário à generalização do método. Assim, o capítulo “2 Aspectos gerais da perícia criminal no ordenamento jurídico brasileiro e aplicação do Teorema de Bayes para a apresentação de seus resultados” foi escrito em termos simplificados, com exemplos práticos, usuais e de resultados inicialmente contraintuitivos que são muitas vezes desconhecidos aos públicos leigo e mesmo técnico, mas sem deixar de detalhar e formalizar as definições empregadas em sua dedução. Esse capítulo visa permitir que uma ampla gama de peritos criminais, com formações superiores nos mais diversos campos do conhecimento, possam partir, desde pouco ou nenhum conhecimento sobre o Teorema de Bayes, e tenham condições de reconhecer, interpretar e aplicar essa técnica em suas análises periciais.

Referências Bibliográficas

- ALI, T.; SPREEUWERS, L.; VELDHUIS, R. Forensic face recognition: A survey. **Face Recognition: Methods, Applications and Technology**, n. February 2014, p. 187–200, 2012.
- ALMEIDA, D. C. de. Atribuição de autoria textual com propósitos forenses : panorama e proposta de análise. **ReVEL**, v. 12, p. 148–186, 2014.
- BEHLAU, M. et al. **Voz - O Livro do Especialista**. Rio de Janeiro: Livraria e Editora Revinter Ltda, 2001. v. 1. ISBN 85-7309-525-3.
- BELLET, A.; HABRARD, A.; SEBBAN, M. A Survey on Metric Learning for Feature Vectors and Structured Data. **CoRR**, abs/1306.6, p. 57, 2013. ISSN 00401951. Disponível em: <<http://arxiv.org/abs/1306.6709>>.
- BERG, J. van den. Myoelastic-Aerodynamic Theory of Voice Production. **Journal of Speech, Language, and Hearing Research**, v. 1, p. 227–244, September 1958.
- BHATTI, A. Cognitive bias in clinical practice – nurturing healthy skepticism among medical students. **Advances in Medical Education and Practice**, Volume 9, p. 235–237, 2018. ISSN 1179-7258.
- BILOTTA, I. et al. How Subtle Bias Infects the Law. **Annual Review of Law and Social Science**, v. 15, p. 227–245, 2019. ISSN 15503585.
- BINONGO, J.; SMITH, M. The application of principal component analysis to stylometry. **Literary and Linguistic Computing**, v. 14, n. 4, p. 445–466, dec 1999. ISSN 0268-1145. Disponível em: <<http://llc.oxfordjournals.org/cgi/doi/10.1093/llc/14.4.445>>.
- BOARETO, M. et al. Supervised Variational Relevance Learning, an analytic geometric feature selection with applications to omic datasets. **IEEE/ACM Transactions on Computational Biology and Bioinformatics**, v. 12, n. 3, p. 705–711, 2015. ISSN 15455963. Disponível em: <<https://doi.org/10.1109/TCBB.2014.2377750>>.
- BRASIL. **Decreto Lei nº 3.689, de 3 de outubro de 1941. Código de Processo Penal**. Rio de Janeiro (RJ), 1941. Disponível em: <http://www.planalto.gov.br/ccivil_03/decreto-lei/del3689.htm>. Acesso em: 14 out. 2020.
- BRASIL. **Constituição da República Federativa do Brasil de 1988**. Brasília (DF), 1988. Disponível em: <http://www.planalto.gov.br/ccivil_03/constituicao/constituicao.htm>. Acesso em: 14 out. 2020.
- BRASIL. **Lei nº 13.105, de 16 de março de 2015. Código de Processo Civil**. Brasília (DF), 2015. Disponível em: <http://www.planalto.gov.br/ccivil_03/_ato2015-2018/2015/lei/l13105.htm>. Acesso em: 14 out. 2020.
- BRASIL. **Lei nº 13.964, de 24 de dezembro de 2019. Pacote Anticrime**. Brasília (DF), 2019. Disponível em: <http://www.planalto.gov.br/ccivil_03/_ato2019-2022/2019/lei/L13964.htm>. Acesso em: 14 out. 2020.

- BURROWS, J. Delta': a Measure of Stylistic Difference and a guide to likely Authorship. **Literary and Linguistic Computing**, v. 17, n. 3, 2002. Disponível em: <<http://llc.oxfordjournals.org/content/17/3/267.short>>.
- CAMPBELL, U. **Suzane: assassina e manipuladora**. 1. ed. São Paulo: Matrix, 2020. 280 p. ISBN 978-85-8230-617-8.
- CHARNOV, E. L. Optimal foraging, the Marginal Value Theorem. **Theoretical Population Biology**, v. 9, p. 129–136, 1976. Disponível em: <<https://www.sciencedirect.com/science/article/pii/004058097690040X>>.
- CHARTRAND, J. P.; PERETZ, I.; BELIN, P. Auditory recognition expertise and domain specificity. **Brain Research**, v. 1220, p. 191–198, 2008. ISSN 00068993.
- CHEN, Y. H. et al. Can knowledge of election results change recall of our predictions? Neural correlates of political hindsight bias. **PLoS ONE**, v. 14, n. 10, p. 1–12, 2019. ISSN 19326203. Disponível em: <<http://dx.doi.org/10.1371/journal.pone.0220690>>.
- CHRÉTIEN, C. D. Review: Who Was Junius? by Alvar Ellegård; A Statistical Method for Determining Authorship: The Junius Letters, 1769-1772 by Alvar Ellegård. **Language (Linguistic Society of America)**, v. 40, n. Jan - Mar, p. 85–90, 1964.
- COLE, S. A.; DIOSO-VILLA, R. Investigating the 'CSI effect' effect: Media and litigation crisis in criminal law. **Stanford Law Review**, v. 61, n. 6, p. 1335–1373, 2009. ISSN 00389765.
- CONTRATADO por Temer, Molina diz que áudio não pode ser usado como prova. **Folha de São Paulo**, 2017. Disponível em: <<https://www1.folha.uol.com.br/poder/2017/05/1886381-contratado-por-temer-molina-diz-que-audio-de-batista-e-prova-impresstavel.shtml>>. Acesso em: 28 mai. 2017.
- COSTA, E. J. d. F. **Levando a imparcialidade a sério: proposta de um modelo interseccional entre direito processual, economia e psicologia**. 187 p. Tese (Doutorado) — Pontifícia Universidade Católica de São Paulo, 2016. Disponível em: <<https://tede.pucsp.br/bitstream/handle/6986/1/EduardoJosedaFonsecaCosta.pdf>>.
- COSTA, E. J. d. F. **Finitude compreensiva do juiz é um dos desafios do Direito Processual**. 2017. Disponível em: <<https://www.conjur.com.br/2017-jan-28/eduardo-costa-finitude-compreensiva-juiz-desafio-direito-processual>>. Acesso em: 25 dez. 2020.
- DAVIS, S. B.; MERMELSTEIN, P. Comparison of Parametric Representations for Monosyllabic Word Recognition in Continuously Spoken Sentences. **IEEE TRANSACTIONS ON ACOUSTICS, SPEECH, AND SIGNAL PROCESSING**, ASSP-28, N, n. 4, p. 357–365, 1980. Disponível em: <<http://dox.doi.org/10.1109/TASSP.1980.1163420>>.
- DEKING, F. M. et al. **A Modern Introduction to Probability and Statistics - Understanding Why and How**. 1. ed. Londres: Springer, 2005. 486 p.
- DWORKIN, R. **O império do direito**. 2. ed. São Paulo: Martins Fontes, 2007. Trad.: Jefferson Luiz Camargo.
- ELLEGÅRD, A. **A Statistical Method for Determining Authorship: The Junius Letters, 1769-1772**. Stockholm, Sweden: Almqvist and Wiksell, 1962.
- ERIKSSON, A. Aural/Acoustic vs. Automatic Methods in Forensic Phonetic Case

Work. In: NEUSTEIN, A.; PATIL, H. A. (Ed.). **Forensic Speaker Recognition: Law Enforcement and Counter-Terrorism**. New York, NY: Springer-Verlag, 2012. cap. 3, p. 41–69. ISBN 9781461402633. Disponível em: <<https://link.springer.com/chapter/10.1007/978-1-4614-0263-3>>.

EUROPEAN NETWORK OF FORENSIC SCIENCE INSTITUTES. **ENFSI: 73 members in 39 countries**. Wiesbaden/Alemanha, 2020. Disponível em: <<https://enfsi.eu/about-enfsi/members/>>. Acesso em: 25 nov. 2020.

FEYNMAN, R. P.; LEIGHTON, R. B.; SANDS, M. L. **The Feynman Lectures on Physics - The New Millennium Edition - Mainly mechanics, radiation, and heat**. [S.l.]: Addison–Wesley, 2010. v. 1. (The Feynman Lectures on Physics, v. 1).

FIGUEIREDO, D. D. de; SAMPAIO, D. A cadeia de custódia na produção probatória penal. In: **PRIMEIRAS IMPRESSÕES SOBRE A LEI 13.964/2019 (PACOTE “ANTICRIME”): A VISÃO DA DEFENSORIA PÚBLICA**. Rio de Janeiro: DEFENSORIA PÚBLICA DO ESTADO DO RIO DE JANEIRO, 2020. cap. 2, p. 28–58. ISBN 978-65-86548-03-7. Disponível em: <<http://www.defensoria.rj.def.br/uploads/arquivos/3989eb7a1ab24ba28dacaec70f32182.pdf>>.

FISCHHOFF, B. Hindsight not equal to Foresight: The effect of outcome knowledge on judgment under uncertainty. **Journal of Experimental Psychology: Human Perception and Performance**, v. 1, n. 3, p. 288–299, 1975. ISSN 00961523.

FISCHHOFF, B.; BEYTH, R. I knew it would happen. Remembered probabilities of once-future things. **Organizational Behavior and Human Performance**, v. 13, n. 1, p. 1–16, 1975. ISSN 00305073.

FLETCHER, H. Loudness, Masking and Their Relation to the Hearing Process and the Problem of Noise Measurement. **The Journal of the Acoustical Society of America**, v. 9, n. 4, p. 275–293, 1938. ISSN NA.

FLETCHER, H.; MUNSON, W. A. Relation between loudness and masking. **The Journal of the Acoustical Society of America**, v. 9, n. 1, p. 1–10, 1937. ISSN 00014966.

FRASER, H. The reliability of voice recognition by ‘ear witnesses’: An overview of research findings. **Language and Law / Linguagem e Direito**, v. 6, n. 2, p. 1–9, 2019. ISSN 2183-3745.

FRAZÃO, F. Perito de temer diz: ‘não garanto a autenticidade da gravação’. **Veja**, 2017. Disponível em: <<https://veja.abril.com.br/politica/perito-de-temer-diz-nao-garanto-a-autenticidade-da-gravacao/>>. Acesso em: 28 mai. 2017.

FUJIOKA, K.; SOUZA, A. de. **NARUHODO #94 – O que é o Teorema de Bayes? (E o que horóscopo tem a ver com isso?)**. 2017. Podcast. Disponível em: <<https://www.b9.com.br/shows/naruhodo/naruhodo-94-o-que-e-o-teorema-de-bayes-e-o-que-horoscopo-tem-a-ver-com-isso/>>. Acesso em: 20 set. 2020.

GIGERENZER, G. et al. **The Empire of Chance: How probability changed science and everyday life**. 1. ed. Cambridge: Cambridge University Press, 1989. 340 p.

GILLIER, R. **O Disfarce da Voz em Fonética Forense**. Dissertação (Mestrado) — Faculdade de Letras da Universidade de Lisboa, 2011.

GILOVICH, T.; MEDVEC, V. H.; CHEN, S. Commission, Omission, and Dissonance

Reduction: Coping with Regret in the "Monty Hall" Problem. **Personality and Social Psychology Bulletin**, v. 21, n. 2, p. 182–190, 1995. ISSN 0146-1672.

GODDARD, C. Thinking across languages and cultures: Six dimensions of variation. **Cognitive Linguistics**, v. 14, n. 2-3, p. 109–140, 2003. ISSN 16133641.

GORDILLO, C. D. A. **Reconhecimento de Voz Contínua Combinando os Atributos MFCC e PNCC com Métodos de Robustez SS, WD, MAP e FRN**. 101 p. Dissertação (Mestrado) — Pontifícia Universidade Católica do Rio de Janeiro, 2013. Disponível em: <<https://doi.org/10.17771/PUCRio.acad.23090>>.

GRANBERG, D. Cross-cultural comparison of responses to the monty hall dilemma. **Social Behavior and Personality**, v. 27, n. 4, p. 431–438, 1999. ISSN 03012212.

GRANBERG, D.; BROWN, T. A. The Monty Hall Dilemma. **Personality and Social Psychology Bulletin**, v. 21, n. 7, p. 711–723, 1995. ISSN 0146-1672.

GURNEY, P.; GURNEY, L. Subsets and Homogeneity: Authorship Attribution in the Scriptoria Historiae Augustae. **Literary and linguistic computing**, v. 13, n. 3, 1998. Disponível em: <<http://llc.oxfordjournals.org/content/13/3/133.short>>.

HAIR JR., J. F. et al. **Multivariate Data Analysis**. 7. ed. [S.l.]: Pearson, 2009.

HANSEN, J. H. L.; HASAN, T. Speaker Recognition by Machines and Humans. **IEEE Signal Processing Magazine**, n. november, p. 74–99, 2015. Disponível em: <<https://doi.org/10.1109/MSP.2015.2462851>>.

HANSEN, J. H. L.; VARADARAJAN, V. Analysis and compensation of lombard speech across noise type and levels with application to in-set/out-of-set speaker recognition. **IEEE Transactions on Audio, Speech and Language Processing**, v. 17, n. 2, p. 366–378, 2009. ISSN 15587916.

HARLEY, E. M. Hindsight bias in legal decision making. **Social Cognition**, v. 25, n. 1, p. 48–63, 2007. ISSN 0278016X.

HASELTON, M. G.; NETTLE, D.; ANDREWS, P. W. The Evolution of Cognitive Bias. **The Handbook of Evolutionary Psychology**, p. 724–746, 2015.

HERBRANSON, W.; SCHROEDER, J. Are Birds Smarter Than Mathematicians? **Computational Psychology**, v. 124, n. 1, p. 1–13, 2010. Disponível em: <<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3086893/>>.

HERBRANSON, W. T. Pigeons, Humans, and the Monty Hall Dilemma. **Current Directions in Psychological Science**, v. 21, n. 5, p. 297–301, 2012. ISSN 09637214.

HOFFRAGE, U.; HERTWIG, R.; GIGERENZER, G. Hindsight Bias: A By-Product of Knowledge Updating? **Journal of Experimental Psychology: Learning Memory and Cognition**, v. 26, n. 3, p. 566–581, 2000. ISSN 02787393.

HOIJTINK, H.; KLUGKIST, I.; BOELEN, P. A. (Ed.). **Bayesian Evaluation of Informative Hypotheses**. New York, USA: Springer Science+Business Media, LLC, 2008. ISBN 9780387096117.

HOLMES, D. I. Authorship attribution. **Computers and Humanities**, v. 28, p. 87–106, 1994.

HUANG, X.; ACERO, A.; HON, H. wuen. **Spoken Language Processing - a guide to theory, algorithm, and system development**. [S.l.]: Prentice Hall PTR, 2001.

IRIBURE, JR., H. d. C.; ROSA, M. d. S. Relativização da presunção de inocência sobre o prisma da integridade e coerência de Ronald Dworkin. **Revista de Direito Penal, Processo Penal e Constituição**, v. 4, n. 1, p. 45–67, 2018.

JOLLIFFE, I. T. **Principal Component Analysis**. 2. ed. New York: Springer, 2002. (Springer series in statistics).

KAHNEMAN, D.; KLEIN, G. Conditions for Intuitive Expertise: A Failure to Disagree. **American Psychologist**, v. 64, n. 6, p. 515–526, 2009. ISSN 0003066X.

KAWAMURA, T. Interpretação de um teste sob a visão epidemiológica: eficiência de um teste. **Arquivos Brasileiros de Cardiologia**, scielo, v. 79, p. 437 – 441, 10 2002. ISSN 0066-782X. Disponível em: <<https://doi.org/10.1590/S0066-782X2002001300015>>.

KUKUCKA, J. et al. Cognitive Bias and Blindness: A Global Survey of Forensic Science Examiners. **Journal of Applied Research in Memory and Cognition**, Society for Applied Research in Memory and Cognition, v. 6, n. 4, p. 452–459, 2017. ISSN 22113681. Disponível em: <<http://dx.doi.org/10.1016/j.jarmac.2017.09.001>>.

KULIS, B. Metric Learning: A Survey. **Foundations and Trends® in Machine Learning**, v. 5, n. 4, p. 287–364, 2013. ISSN 1935-8237. Disponível em: <<http://ieeexplore.ieee.org/xpl/articleDetails.jsp?arnumber=8186753>>.

LANGER, E. J. The illusion of control. **Journal of Personality and Social Psychology**, v. 32, n. 2, p. 311–328, 1975. ISSN 00223514.

LIMA NETO, E. G.; ALBANO, D. F. L. Influência da Prova Pericial em Sentenças Judiciais de Casos Definidos na Lei de Drogas em Aracaju entre 2012 e 2018. **Revista Brasileira de Criminalística**, v. 9, n. 2, p. 113, 2020. ISSN 2237-9223.

MAHER, R. Audio forensic examination. **Signal Processing Magazine, IEEE**, n. March, p. 84–94, 2009. Disponível em: <http://ieeexplore.ieee.org/xpls/abs/_all.jsp?arnumber=4806>.

MALTA, F. et al. **Scicast # 377: Teorema de Bayes**. 2020. Podcast. Disponível em: <<https://www.deviante.com.br/podcasts/scicast-377/>>. Acesso em: 20 set. 2020.

MASTERSON, M. Your voice is your password. **Speech Technology Magazine**, p. 18–21, 2012.

MORRISON, G. et al. Forensic database of voice recordings of 500+ Australian English speakers. 2015. Disponível em: <<http://databases.forensic-voice-comparison.net/>>.

MORRISON, G. S. Forensic voice comparison and the paradigm shift. **Science & justice : journal of the Forensic Science Society**, Forensic Science Society, v. 49, n. 4, p. 298–308, dec 2009. ISSN 1355-0306. Disponível em: <<http://dx.doi.org/10.1016/j.scijus.2009.09.002>>.

MORRISON, G. S.; ROSE, P.; ZHANG, C. Protocol for the collection of databases of recordings for forensic-voice-comparison research and practice. **Australian Journal of Forensic Sciences**, v. 44, n. 2, p. 155–167, 2012. ISSN 1834562X. Disponível em: <<http://dx.doi.org/10.1080/00450618.2011.630412>>.

MÜLLER-LYER ILLUSION. **Müller-Lyer illusion – Wikipedia, The Free Encyclopedia**. 2021. Disponível em: <https://en.wikipedia.org/wiki/M%C3%BCller-Lyer_illusion>. Acesso em: 12 jan. 2021.

NASCIMENTO, I. **Uma visão amigável do Teorema de Bayes**. 2017. Disponível em: <<https://lamfo-unb.github.io/2017/08/04/Uma-visao-amigavel-do-Teorema-de-Bayes/>>. Acesso em: 10 set. 2020.

NOLAN, F. **The Phonetic Bases of Speaker Recognition**. [S.l.]: Cambridge University Press, 1983. (Cambridge Studies in Speech Science and Communication).

NUSSENZVEIG, H. M. **Fluidos, Oscilações e Ondas, Calor**. 4. ed. São Paulo: Edgard Blücher, 2002. v. 2. (Curso de Física Básica, v. 2).

OPPENHEIM, A. V.; WILLISKY, A. S.; NAWAB, S. H. **Sinais e Sistemas**. 2. ed. [S.l.]: Pearson Prentice Hall, 2010.

ORGANIZAÇÃO DAS NAÇÕES UNIDAS. **Declaração Universal dos Direitos Humanos**. Paris: Escritório do Alto Comissariado das Nações Unidas para os Direitos Humanos, 1948. Disponível em: <<https://www.ohchr.org/EN/UDHR/Pages/Language.aspx?LangID=por>>. Acesso em: 15 dez. 2020.

ORGANIZAÇÃO DOS ESTADOS AMERICANOS. **Convenção Americana sobre Direitos Humanos**. San José: Comissão Interamericana de Direitos Humanos, 1969. Assinada na Conferência Especializada Interamericana sobre Direitos Humanos, San José, Costa Rica, em 22 de novembro de 1969. Disponível em: <<https://www.ohchr.org/EN/UDHR/Pages/Language.aspx?LangID=por>>. Acesso em: 15 dez. 2020.

O'SHAUGHNESSY, D. **Speech Communications - Human and Machine**. 2. ed. [S.l.]: IEEE Press, 2000.

O'SULLIVAN, E. D.; SCHOFIELD, S. J. Cognitive bias clinical medicine. **Journal of the Royal College of Physicians of Edinburgh**, v. 48, n. 3, p. 225–232, 2018. ISSN 14782715.

PETRY, A. **Reconhecimento Automático de Locutor Utilizando Medidas de Invariantes Dinâmicas Não-Lineares**. Tese (Doutorado) — Universidade Federal do Rio Grande do Sul, 2002. Disponível em: <<http://www.lume.ufrgs.br/handle/10183/3111>>.

PRONIN, E.; LIN, D. Y.; ROSS, L. The bias blind spot: Perceptions of bias in self versus others. **Personality and Social Psychology Bulletin**, v. 28, n. 3, p. 369–381, 2002. ISSN 01461672.

RGB. **RGB – Wikipedia, The Free Encyclopedia**. 2021. Disponível em: <<https://pt.wikipedia.org/wiki/RGB>>. Acesso em: 06 mar. 2021.

ROSE, P. **Forensic Speaker Identification**. [S.l.]: Taylor & Francis, 2002. (Taylor & Francis Forensic Science Series).

ROSS, A.; JAIN, A. Human recognition using biometrics: an overview. **Annales Des Télécommunications**, p. 11–35, 2007. Disponível em: <<http://link.springer.com/article/10.1007/BF03253248>>.

RUDMAN, J. The state of authorship attribution studies: Some problems and solutions. **Computers and the Humanities**, p. 351–365, 1997. Disponível em: <<http://link.springer.com/article/10.1023/A:1001018624850>>.

RUDMAN, J. The state of authorship attribution studies: Some problems and solutions. **Computers and the Humanities**, v. 93, n. 3, p. 259–274, may 1997. ISSN

0013-838X. Disponível em: <<http://www.tandfonline.com/doi/abs/10.1080/0013838X.2012.668785><http://link.springer.com/article/10.1023/A:1001018624850>>.

SÃO PAULO. **Resolução SSP-142 (GS-261/13)**. São Paulo (SP), 2014. Disponível em: <https://www.imprensaoficial.com.br/DO/BuscaDO2001Documento_11_4.aspx?link=/2014/executivo%2520secao%2520i/outubro/10/pag_0010_3OCQUCDI085O7e4RDATDGLBA0R0.pdf&pagina=10&data=10/10/2014&caderno=Executivo+l&paginaordenacao=100010>. Acesso em: 27 jan. 2020.

SAKS, M. J.; KOEHLER, J. J. The coming paradigm shift in forensic identification science. **Science**, v. 309, n. 5736, p. 892–896, 2005. Disponível em: <link.gale.com/apps/doc/A135247478/AONE?u=capes{&}sid=AONE{&}xid=08>.

SALA, D. A perícia criminal: evidências, profissional perito e nulidade pericial – uma revisão literária. **Revista Brasileira de Criminalística**, v. 7, n. 3, p. 28–31, 2018. ISSN 2237-3691.

SALERNO, J. M.; DIAMOND, S. S. The promise of a cognitive perspective on jury deliberation. **Psychonomic Bulletin and Review**, v. 17, n. 2, p. 174–179, 2010. ISSN 10699384.

SAMPLING (SIGNAL PROCESSING). **Sampling (signal processing) – Wikipedia, The Free Encyclopedia**. 2017. Disponível em: <[https://en.wikipedia.org/wiki/Sampling_\(signal_processing\)](https://en.wikipedia.org/wiki/Sampling_(signal_processing))>. Acesso em: 29 mai. 2017.

SANCHES, A. P. **Palestra: Comunicação Humana Para Uso Forense**. 2012. Disponível em: <www.acadefor.com.br>.

SILVA, E. R.; TABACNIKS, M. H. Supervised Variational Relevance Learning (SUVREL) applied to voice comparison. In: Julien Epps, Joe Wolfe, J. S.; JONES, C. (Ed.). **Proceedings of the Seventeenth Australasian International Conference on Speech Science and Technology**. Sydney, Australia: The Australasian Speech Science & Technology Association Incorporated, 2018. p. 53–56. Disponível em: <<https://bit.ly/3xIFk4C>>.

SILVA, N. L. S. e. **OS VIESES DE COGNIÇÃO E O PROCESSO JURISDICIONAL DEMOCRÁTICO: Um estudo sobre a mitigação de seus efeitos e o debiasing**. 137 p. Dissertação (Mestrado) — Pontifícia Universidade Católica de Minas Gerais, 2018.

SILVA, W. C. da. **O princípio da presunção de inocência ou não culpabilidade**. 2018. Disponível em: <<https://jus.com.br/artigos/64135/o-principio-da-presuncao-de-inocencia-ou-nao-culpabilidade>>. Acesso em: 06 nov. 2020.

SLOOTEN, K. Likelihood ratio distributions and the (ir)relevance of error rates. **Forensic Science International: Genetics**, Elsevier, v. 44, n. September 2019, p. 102173, 2020. ISSN 18780326. Disponível em: <<https://doi.org/10.1016/j.fsigen.2019.102173>>.

SOARES, F. d. S. O. A autonomia da perícia criminal e a produção de provas face às garantias constitucionais. **Vertentes do Direito**, v. 7, n. 1, p. 231–255, 2020.

SOUSA, L. F. P. de. A valoração da prova pericial. **Revista Portuguesa do Dano Corporal**, n. 27, p. 11–24, 2016. ISSN 1645-0760.

STEIN, B.; LIPKA, N.; PRETTENHOFER, P. Intrinsic plagiarism analysis. **Language Resources and Evaluation**, v. 45, n. 1, p. 63–82, jan 2010. ISSN 1574-020X.

Disponível em: <<http://link.springer.com/10.1007/s10579-010-9115-y>>.

STEVENS, S. S.; VOLKMANN, J. The relation of pitch to frequency: A revised scale. **The American Journal of Psychology**, v. 53, n. 3, p. 329–353, 1940. ISSN 00029556. Disponível em: <<http://www.jstor.org/pss/1417526>>.

STEVENS, S. S.; VOLKMANN, J.; NEWMAN, E. B. A Scale for Measurement of the Psychological Magnitude Pitch. **Journal of the Acoustical Society of America**, v. 8, n. jan 1937, p. 185–190, 1937.

SUN, J. et al. Applying Stylometric Analysis Techniques to Counter Anonymity in Cyberspace. **Journal of Networks**, v. 7, n. 2, p. 259–266, feb 2012. ISSN 1796-2056. Disponível em: <<http://ojs.academypublisher.com/index.php/jnw/article/view/5727>>.

TARONI, F. et al. **Data analysis in forensic science: A bayesian decision perspective**. 1. ed. West Sussex - UK: John Wiley and Sons, Ltd., 2010. v. 88. 367 p. (Statistics in Practice, v. 88). ISBN 9780470665077.

TAVARES, A. A. G.; ANDRADE, O. D. O. Pressupostos Constitucionais da Prova Pericial no Processo Penal. **Revista de Ciências Jurídicas e Sociais**, v. 3, n. 1, p. 7–17, 2013.

WARD, P. et al. Adaptive Skill as the Conditio Sine Qua Non of Expertise. **Journal of Applied Research in Memory and Cognition**, The Society for Applied Research in Memory and Cognition, v. 7, n. 1, p. 35–50, 2018. ISSN 22113681. Disponível em: <<https://doi.org/10.1016/j.jarmac.2018.01.009>>.

WASSERMAN, L. **All of Statistics - A Concise Course in Statistical Inference**. 1. ed. New York: Springer, 2003. 442 p.

WIERZBICKA, A. **Semantics: primes and universals**. 1. ed. Oxford: Oxford University Press, 1996. 500 p.

WILLIS, S. **ENFSI Guideline for Evaluative Reporting in Forensic Science: Strengthening the Evaluation of Forensic Results across Europe (STEOFRAE)**. [S.l.], 2015. 128 p. Disponível em: <<https://enfsi.eu/documents/forensic-guidelines/>>.

WORLDOMETERS.INFO. **COVID-19 CORONAVIRUS PANDEMIC**. Dover, Delaware, U.S.A: [s.n.], 2020. Disponível em: <<https://www.worldometers.info/coronavirus/#countries>>. Acesso em: 25 maio 2020.

ZAPF, P. A. et al. Cognitive bias in forensic mental health assessment: Evaluator beliefs about its nature and scope. **Psychology, Public Policy, and Law**, v. 24, n. 1, p. 1–10, 2018. ISSN 10768971.

ZHENG, R. et al. A framework for authorship identification of online messages: Writing-style features and classification techniques. **Journal of the American ...**, v. 57, n. 3, p. 378–393, 2006. Disponível em: <<http://onlinelibrary.wiley.com/doi/10.1002/asi.20316/full>>.

ZUIKER, A. E. **CSI: Crime Scene Investigation**. New York City: CBS, 2000–2015. Disponível em: <<https://www.cbs.com/shows/csi/>>.

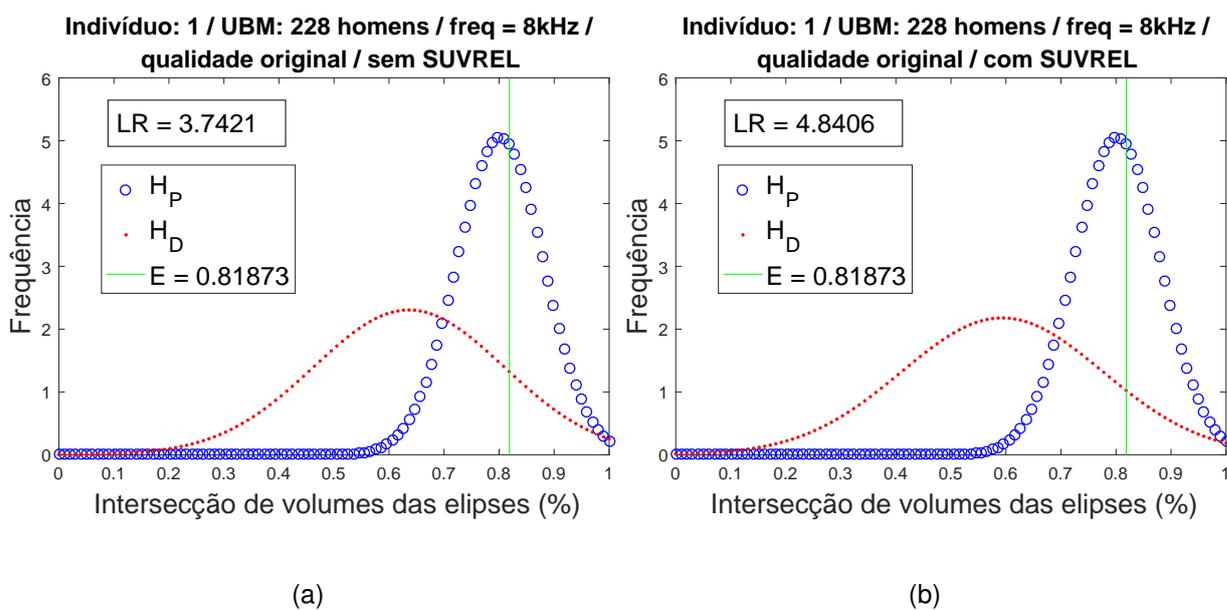
Apêndices

A. Gráficos

Neste capítulo, são apresentados todos os gráficos resultantes dos dados listados nas tabelas 5.6 a 5.9.

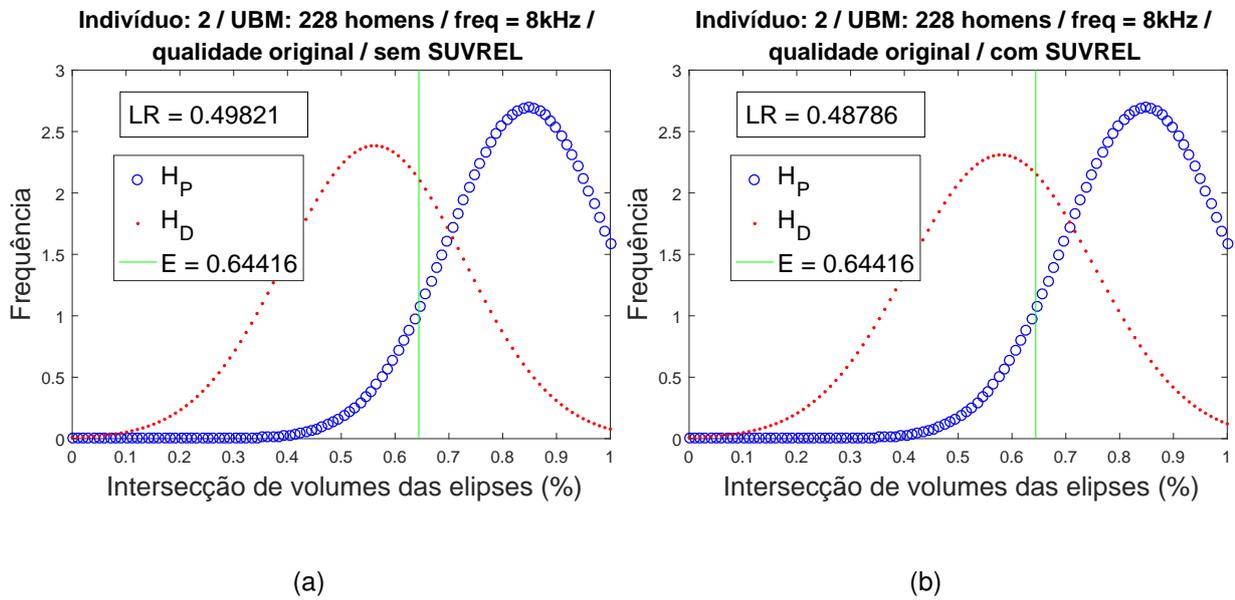
A.1. UBM: 228 homens / frequência: 8kHz / qualidade original

Figura A.1. – Cálculo de LR no cenário apresentado no título de cada gráfico.



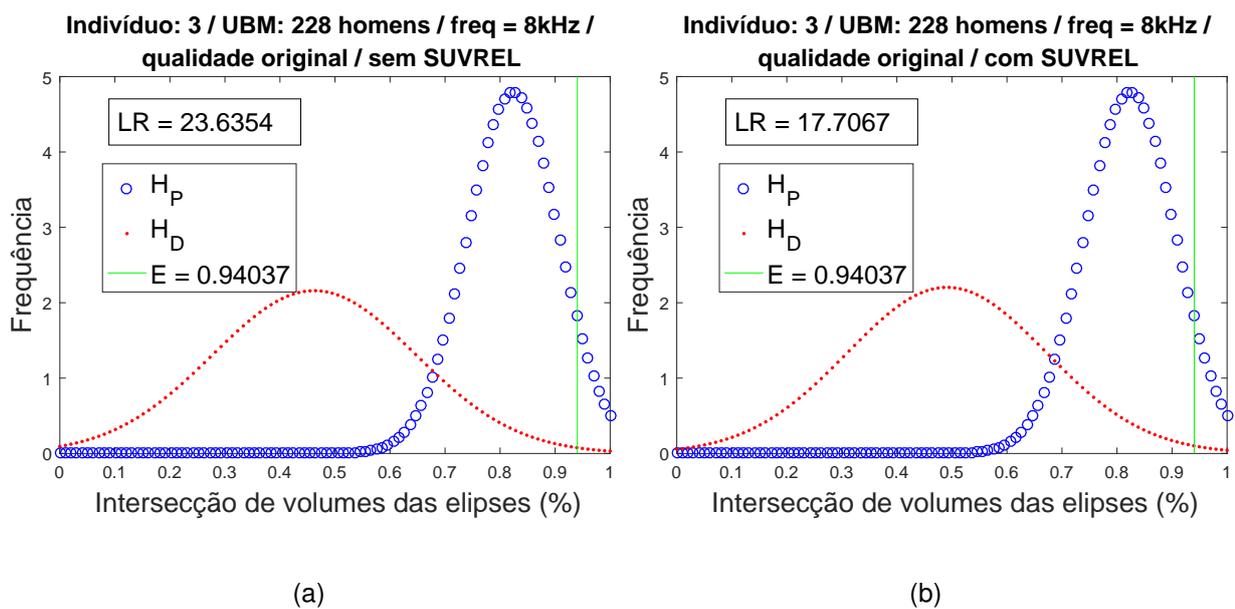
Fonte: produção própria.

Figura A.2. – Cálculo de LR no cenário apresentado no título de cada gráfico.



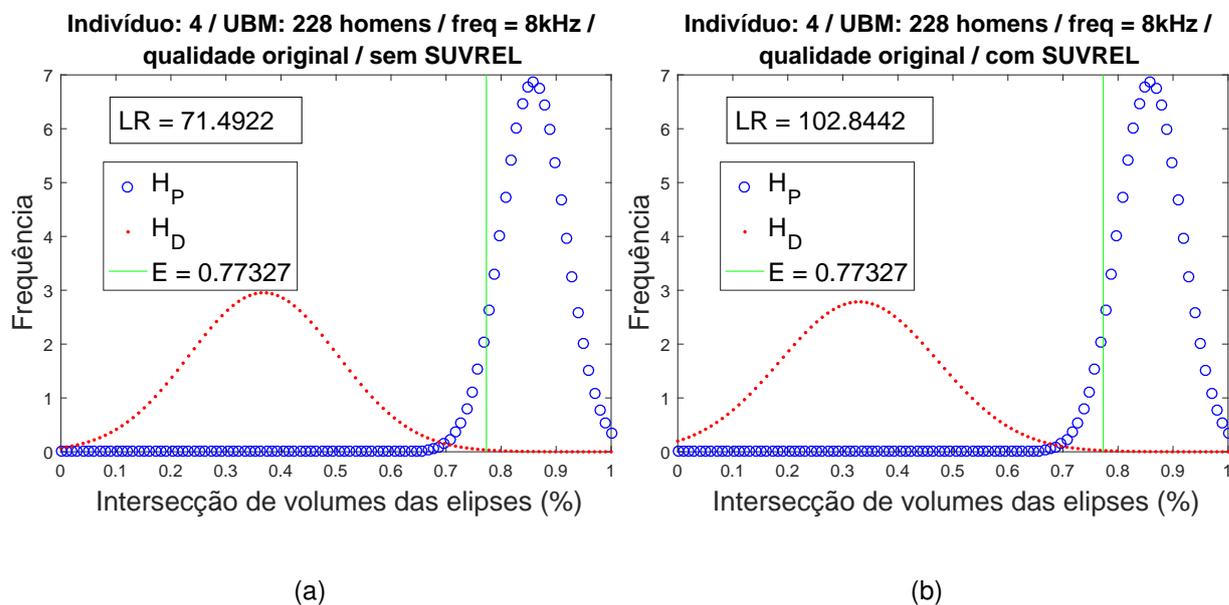
Fonte: produção própria.

Figura A.3. – Cálculo de LR no cenário apresentado no título de cada gráfico.



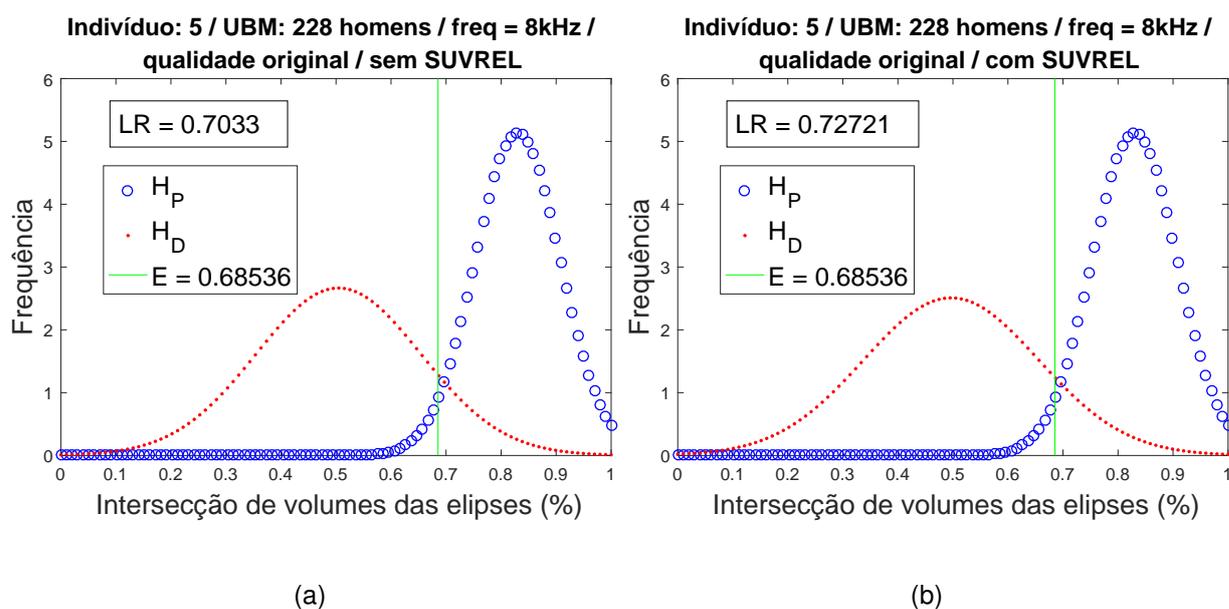
Fonte: produção própria.

Figura A.4. – Cálculo de LR no cenário apresentado no título de cada gráfico.



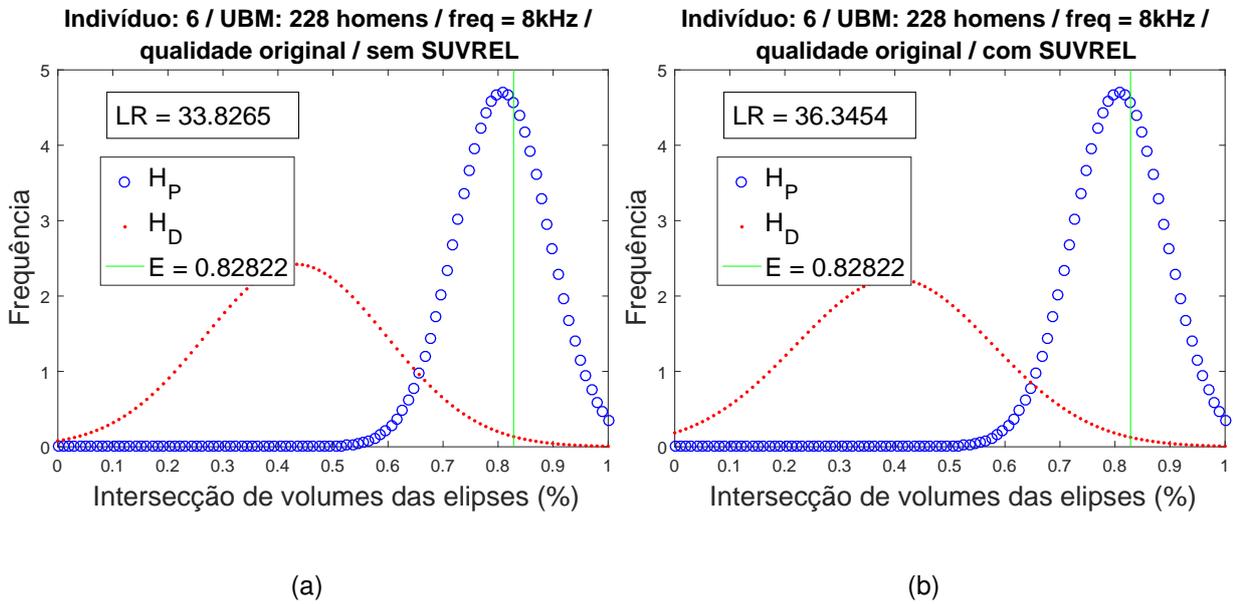
Fonte: produção própria.

Figura A.5. – Cálculo de LR no cenário apresentado no título de cada gráfico.



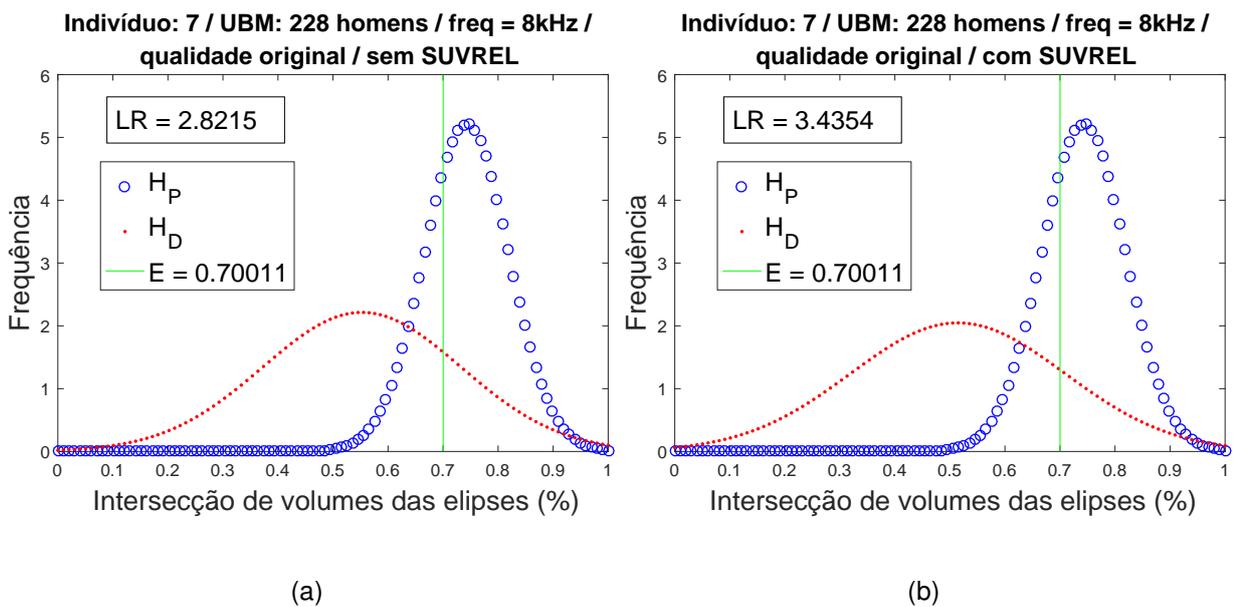
Fonte: produção própria.

Figura A.6. – Cálculo de LR no cenário apresentado no título de cada gráfico.



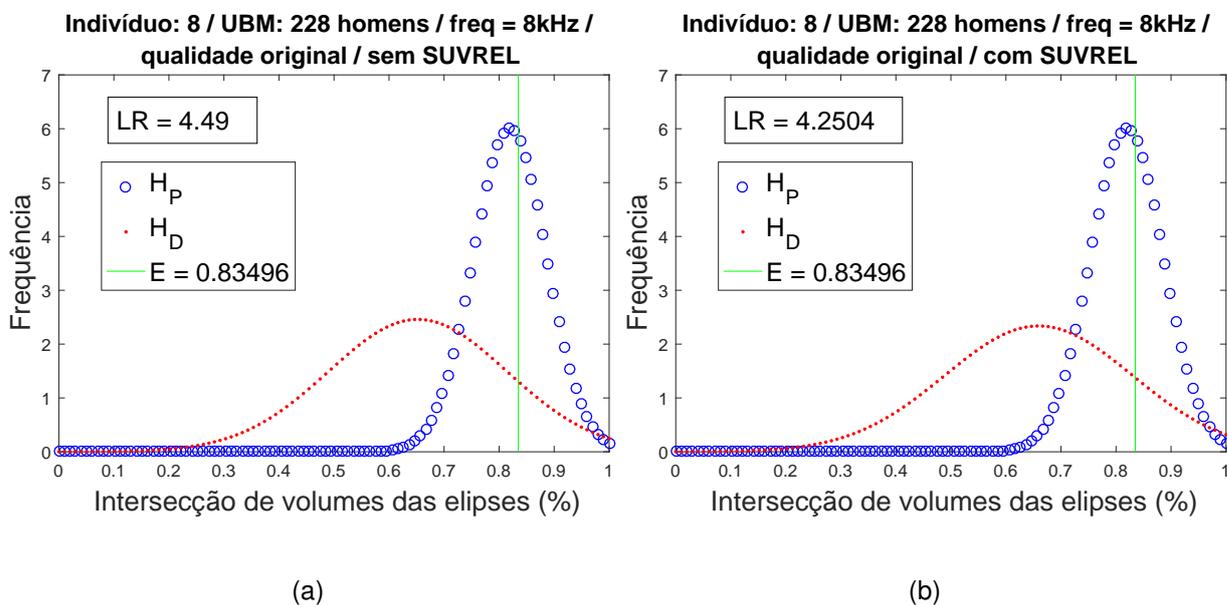
Fonte: produção própria.

Figura A.7. – Cálculo de LR no cenário apresentado no título de cada gráfico.



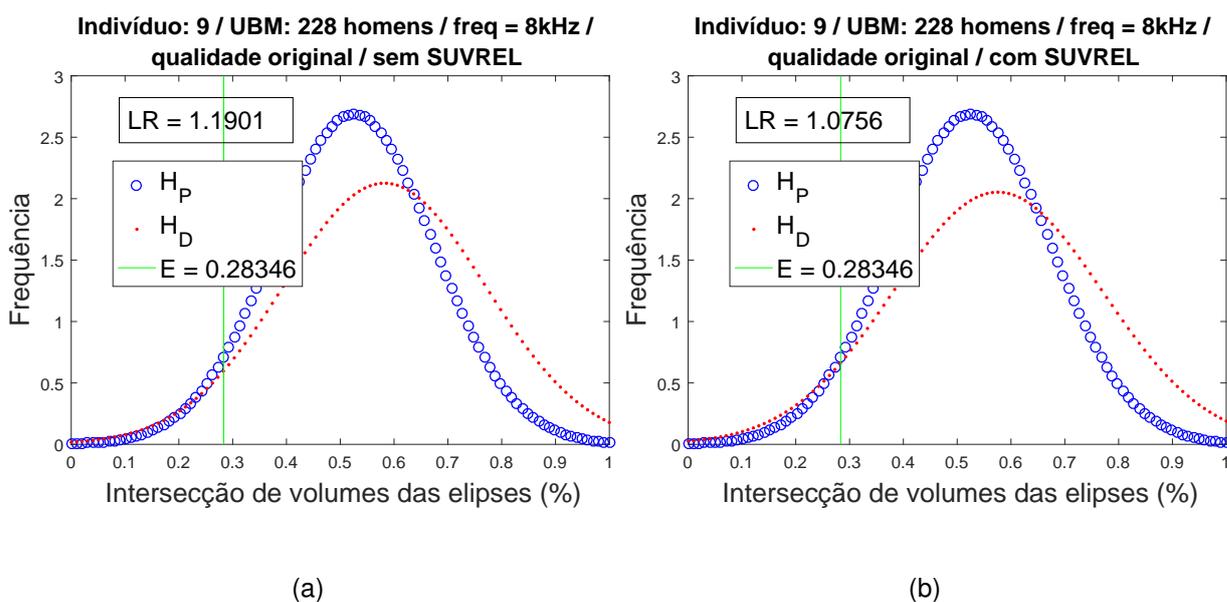
Fonte: produção própria.

Figura A.8. – Cálculo de LR no cenário apresentado no título de cada gráfico.



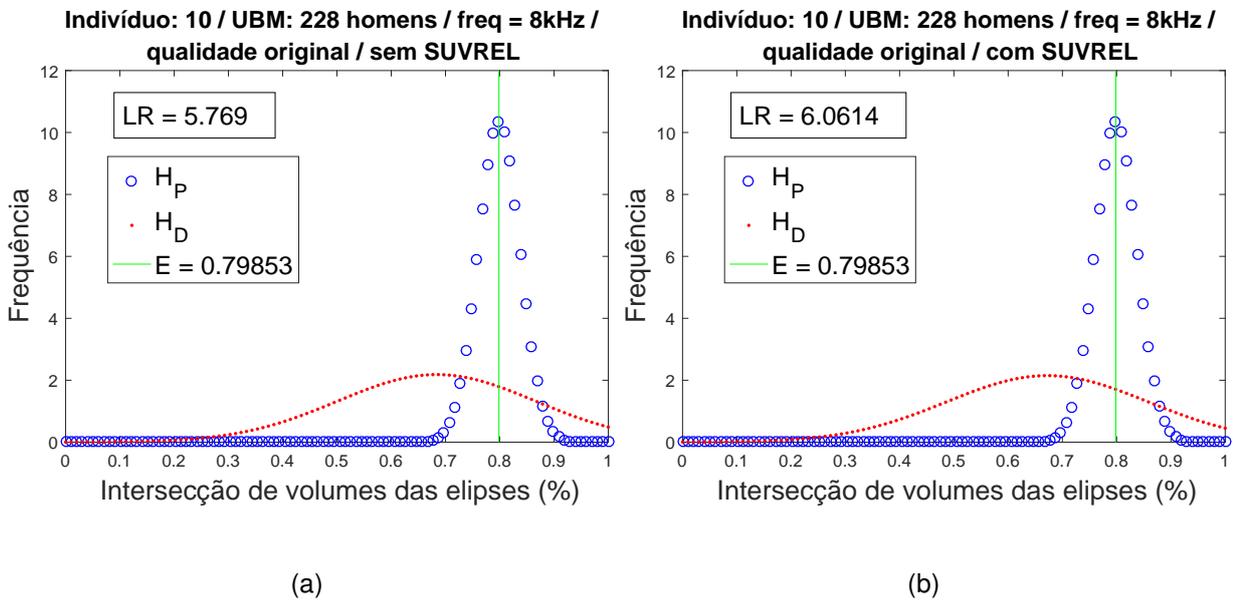
Fonte: produção própria.

Figura A.9. – Cálculo de LR no cenário apresentado no título de cada gráfico.



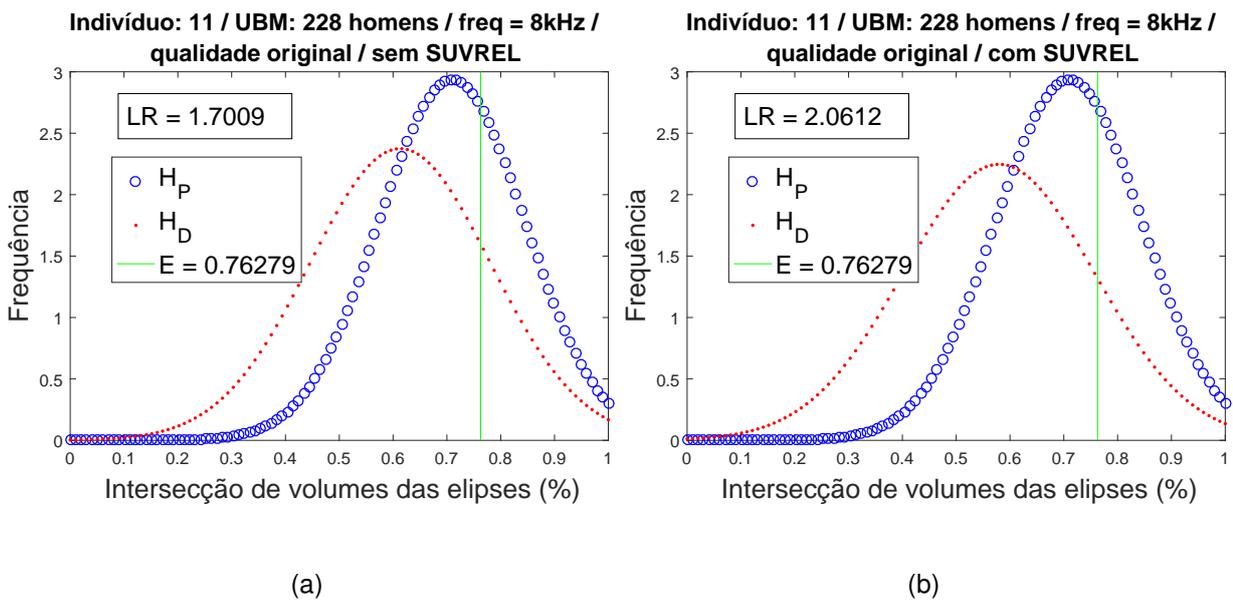
Fonte: produção própria.

Figura A.10. – Cálculo de LR no cenário apresentado no título de cada gráfico.



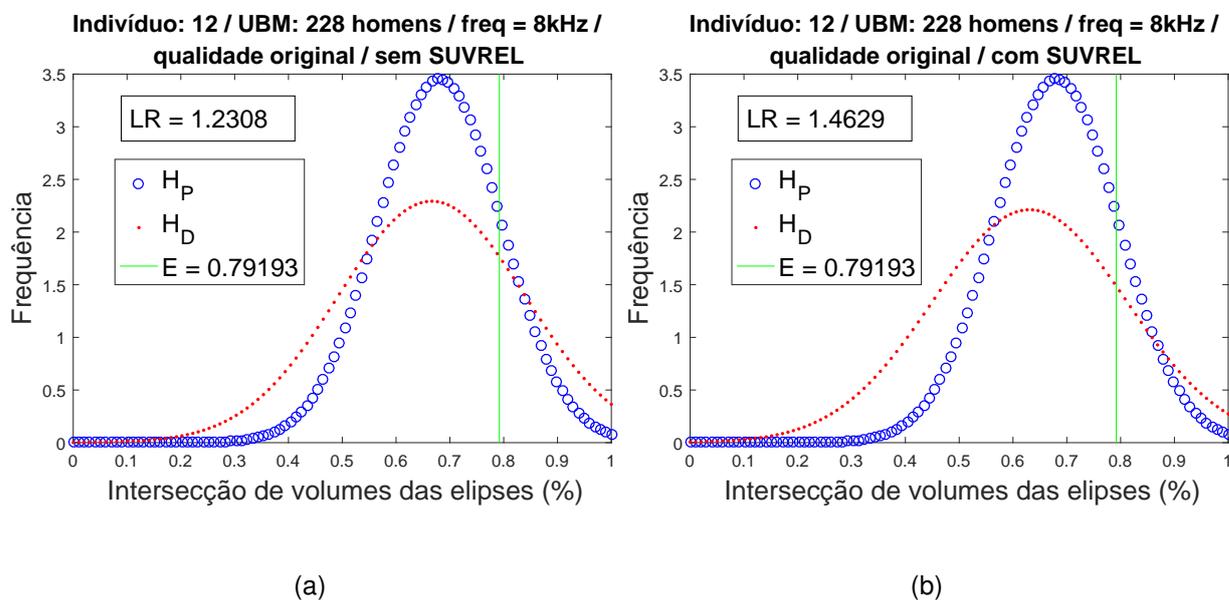
Fonte: produção própria.

Figura A.11. – Cálculo de LR no cenário apresentado no título de cada gráfico.



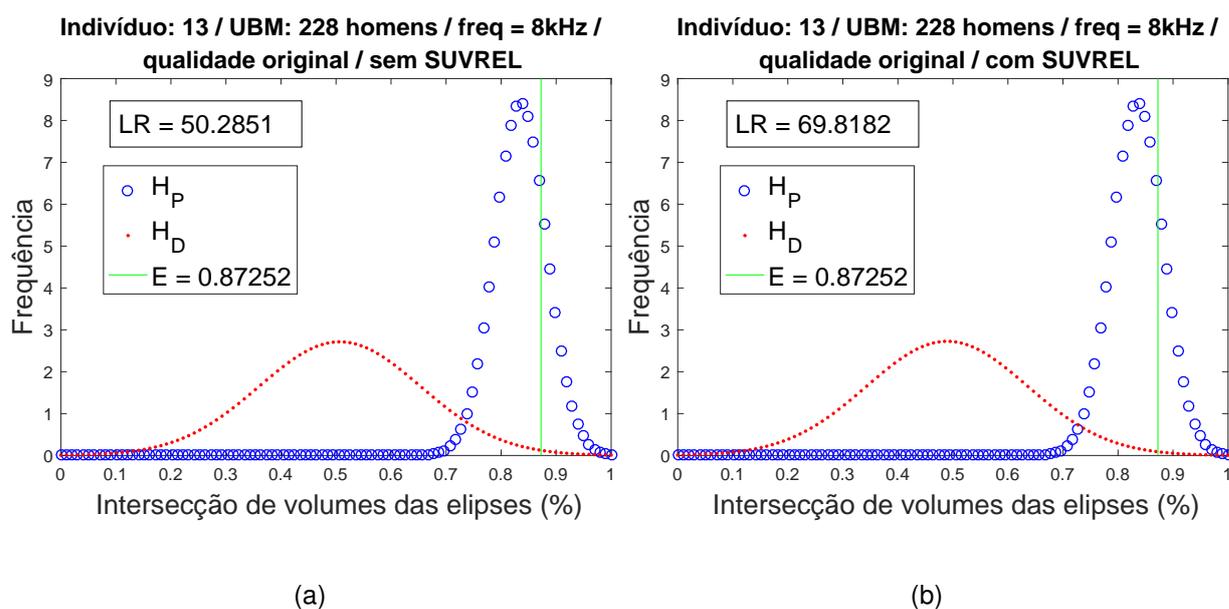
Fonte: produção própria.

Figura A.12. – Cálculo de LR no cenário apresentado no título de cada gráfico.



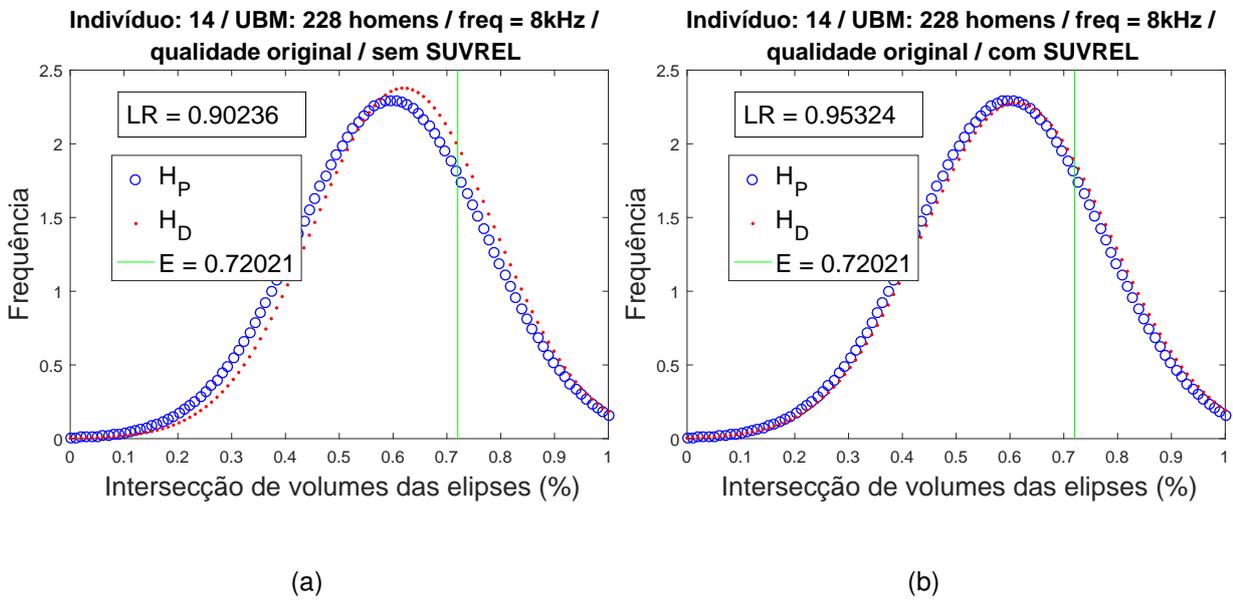
Fonte: produção própria.

Figura A.13. – Cálculo de LR no cenário apresentado no título de cada gráfico.



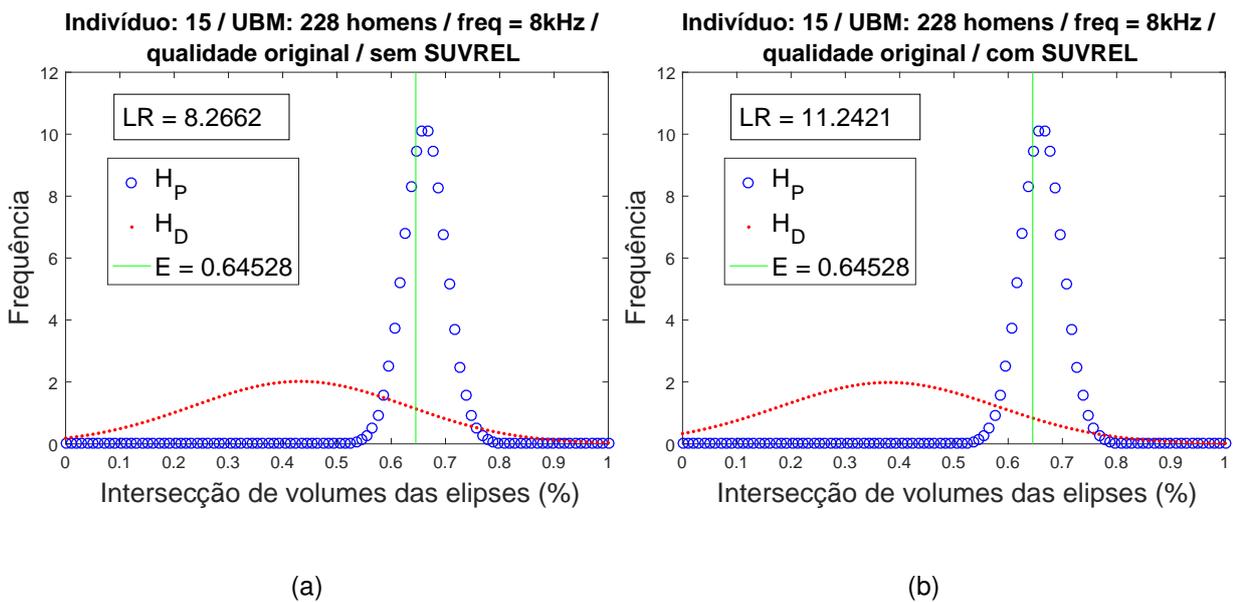
Fonte: produção própria.

Figura A.14. – Cálculo de LR no cenário apresentado no título de cada gráfico.



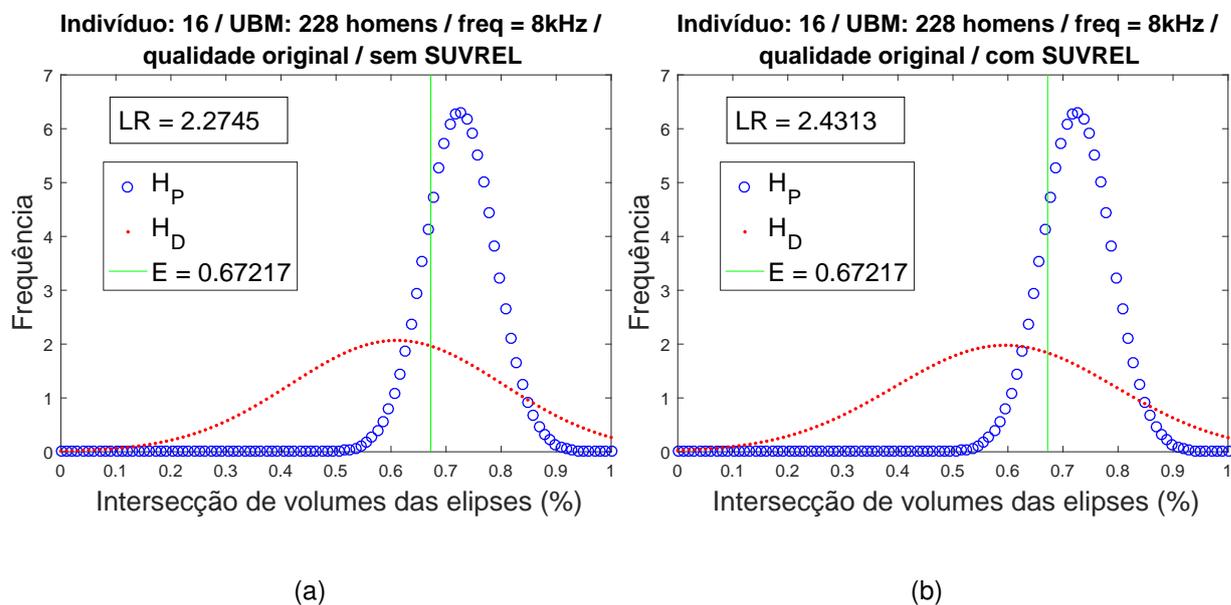
Fonte: produção própria.

Figura A.15. – Cálculo de LR no cenário apresentado no título de cada gráfico.



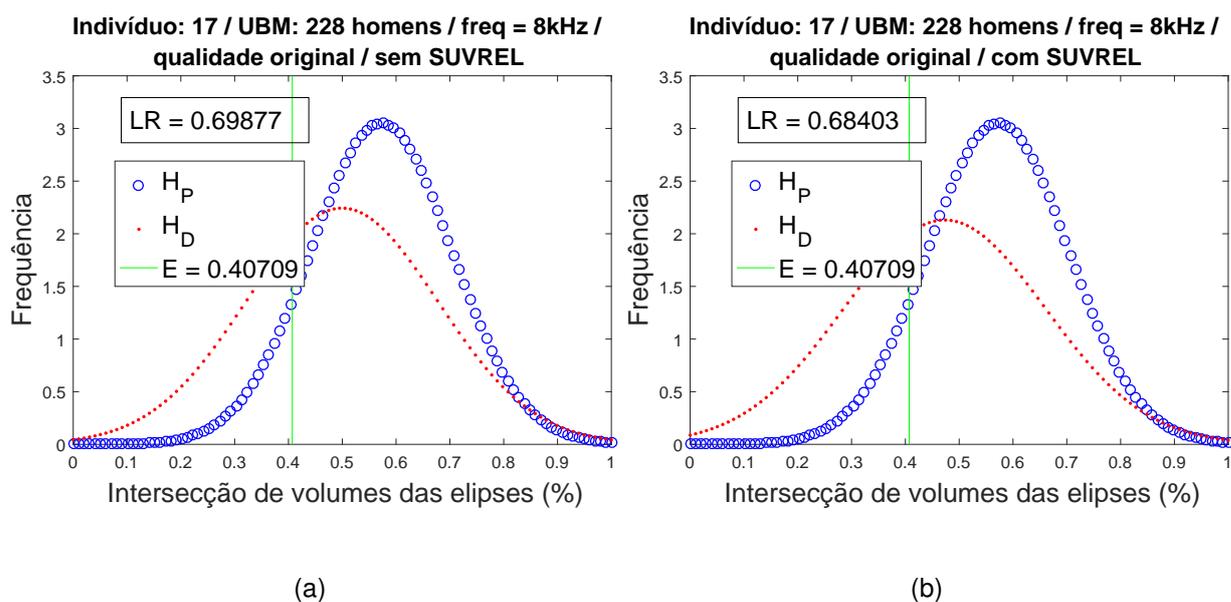
Fonte: produção própria.

Figura A.16. – Cálculo de LR no cenário apresentado no título de cada gráfico.



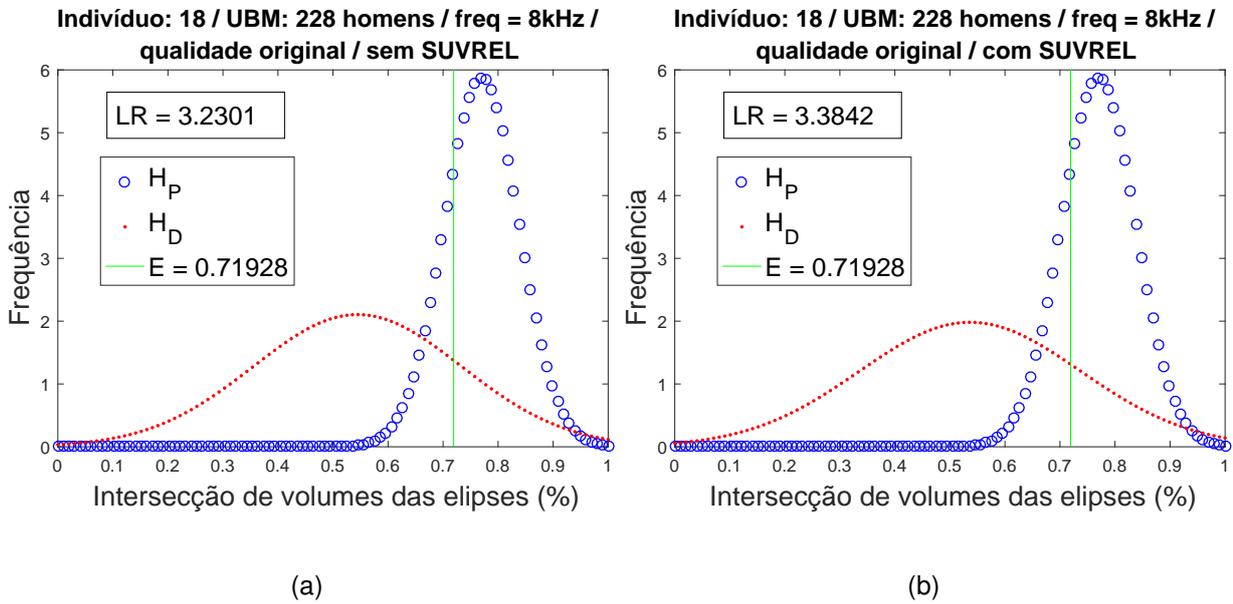
Fonte: produção própria.

Figura A.17. – Cálculo de LR no cenário apresentado no título de cada gráfico.



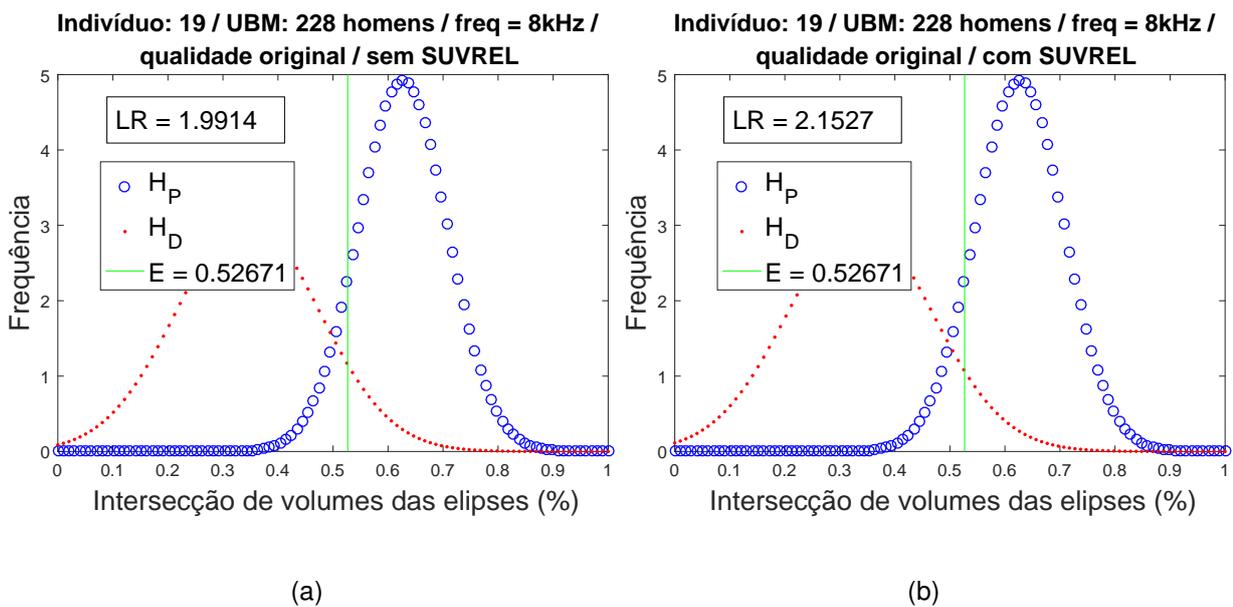
Fonte: produção própria.

Figura A.18. – Cálculo de LR no cenário apresentado no título de cada gráfico.



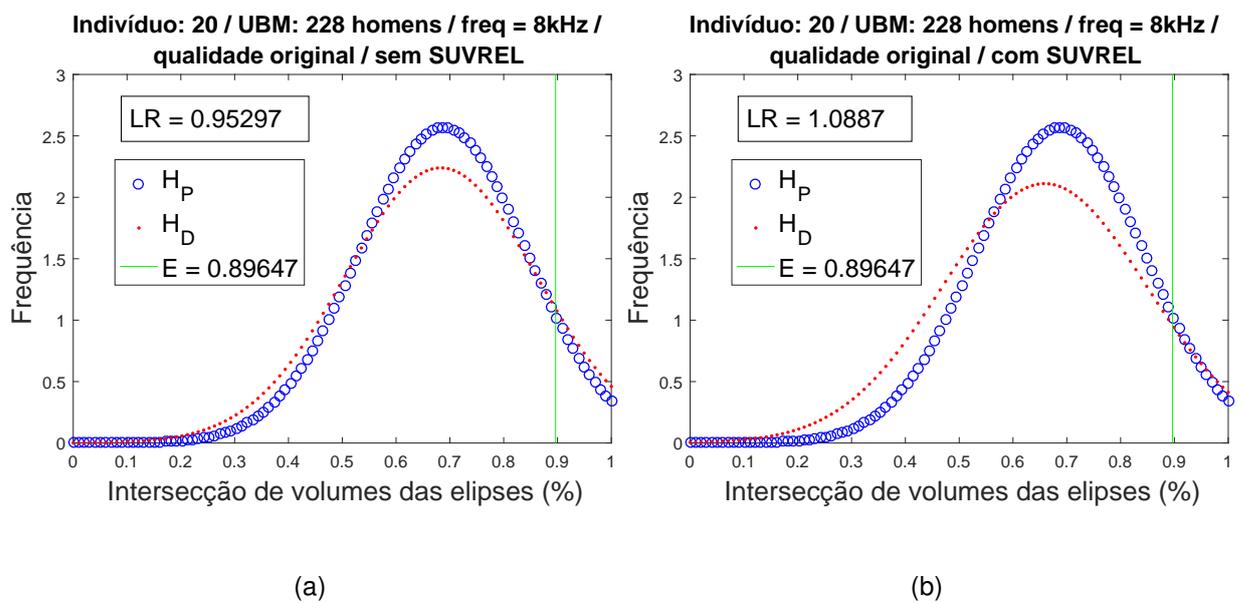
Fonte: produção própria.

Figura A.19. – Cálculo de LR no cenário apresentado no título de cada gráfico.



Fonte: produção própria.

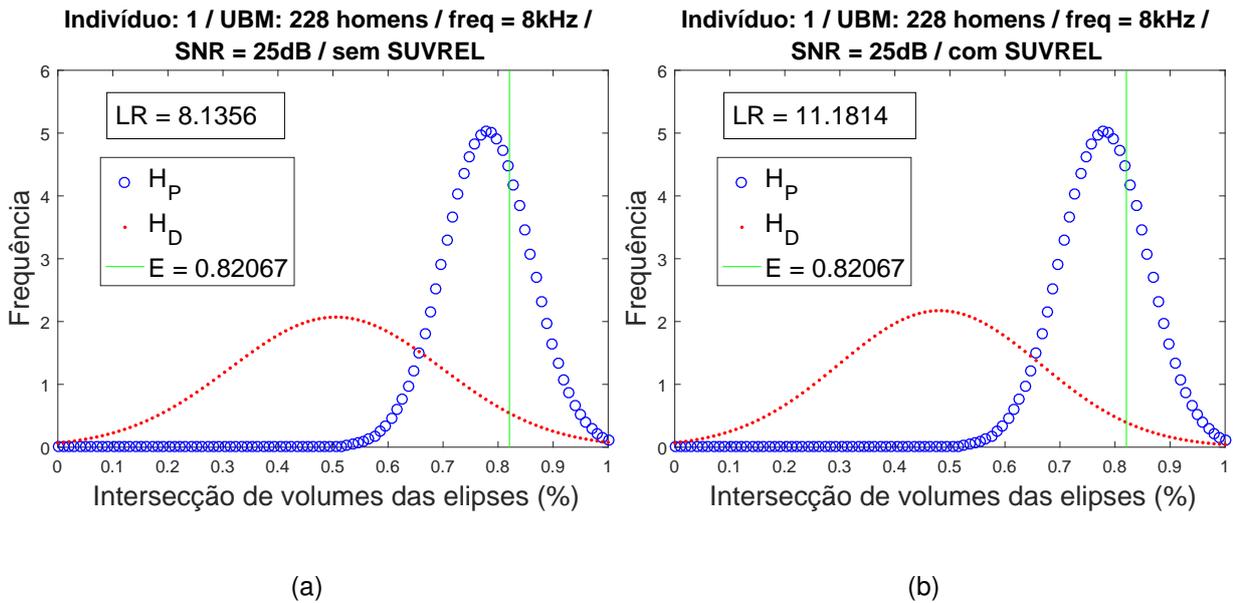
Figura A.20. – Cálculo de LR no cenário apresentado no título de cada gráfico.



Fonte: produção própria.

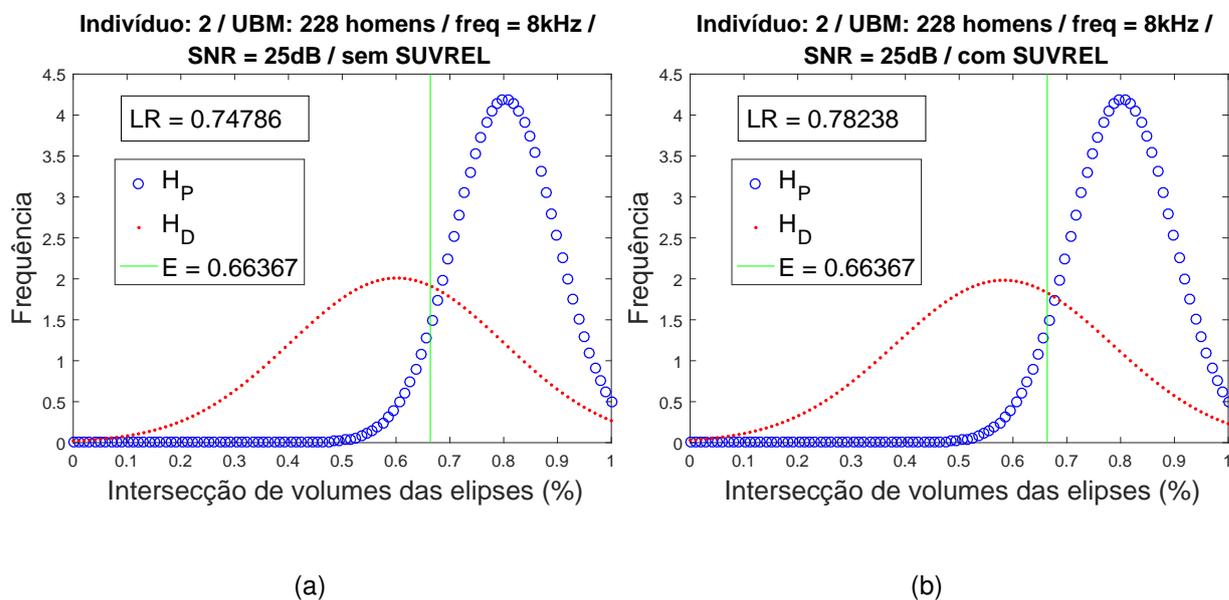
A.2. UBM: 228 homens / frequência: 8kHz / SNR: 25dB

Figura A.21. – Cálculo de LR no cenário apresentado no título de cada gráfico.



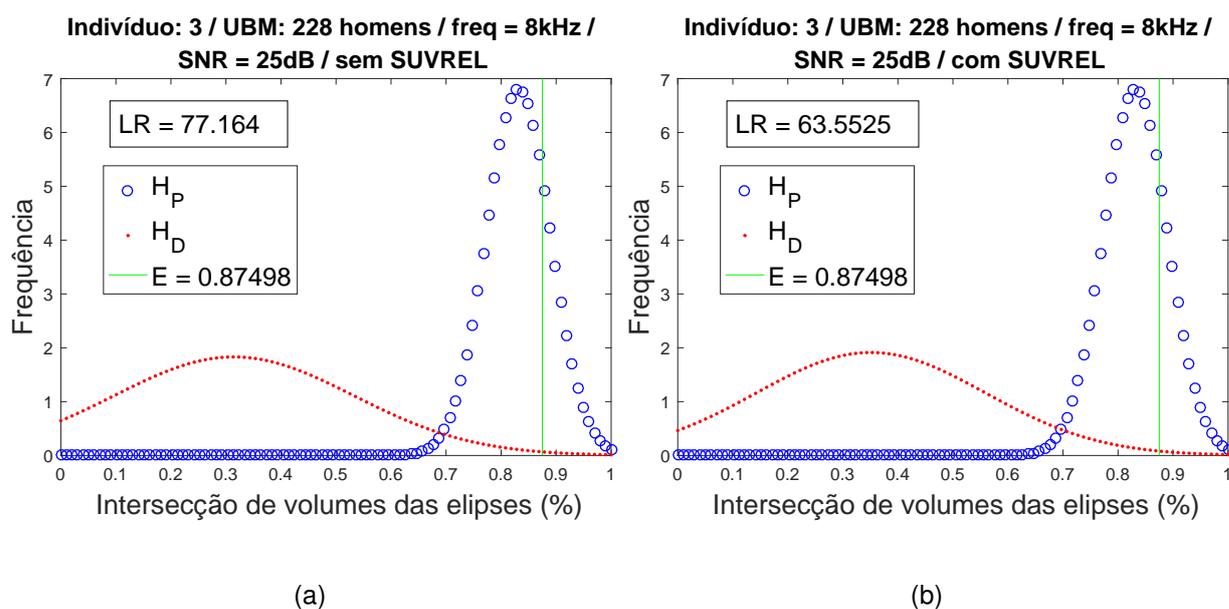
Fonte: produção própria.

Figura A.22. – Cálculo de LR no cenário apresentado no título de cada gráfico.



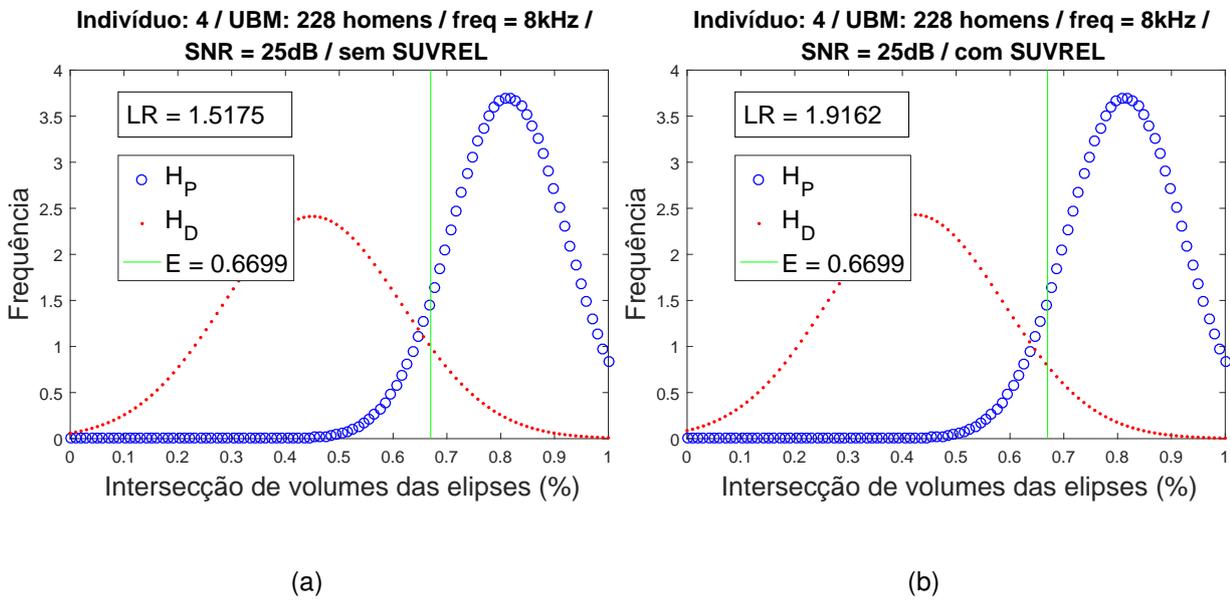
Fonte: produção própria.

Figura A.23. – Cálculo de LR no cenário apresentado no título de cada gráfico.



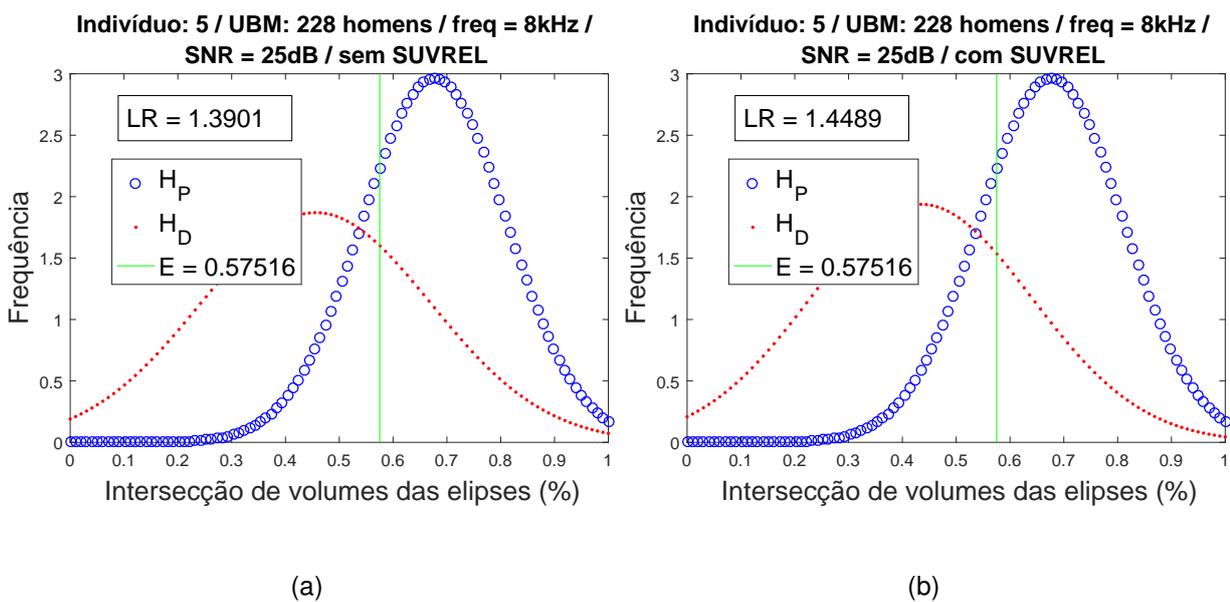
Fonte: produção própria.

Figura A.24. – Cálculo de LR no cenário apresentado no título de cada gráfico.



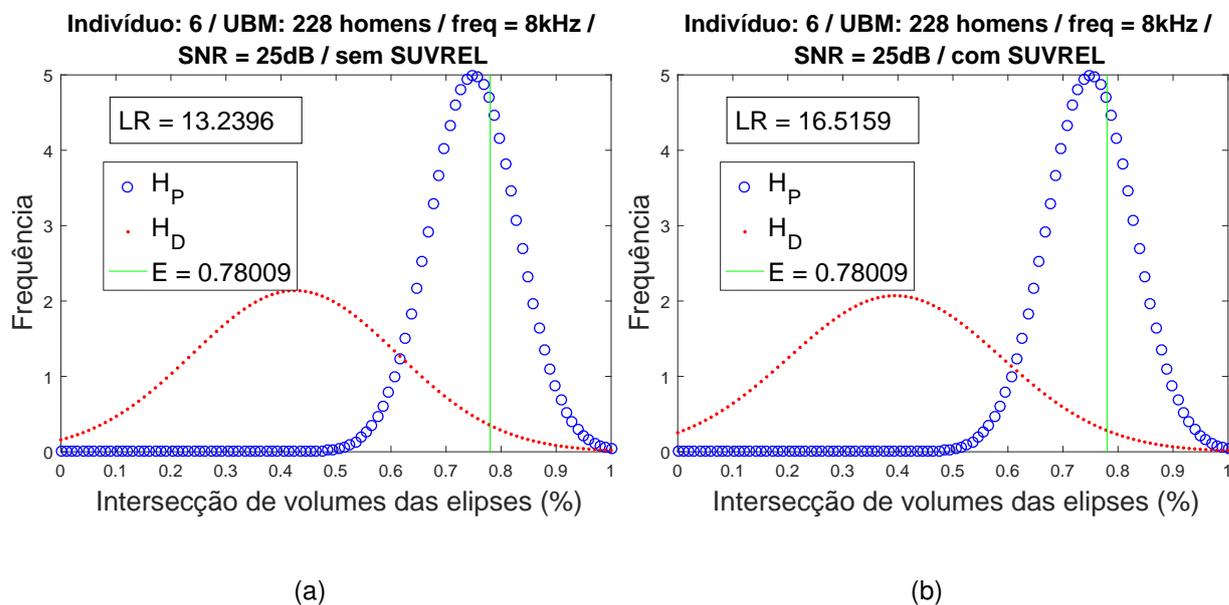
Fonte: produção própria.

Figura A.25. – Cálculo de LR no cenário apresentado no título de cada gráfico.



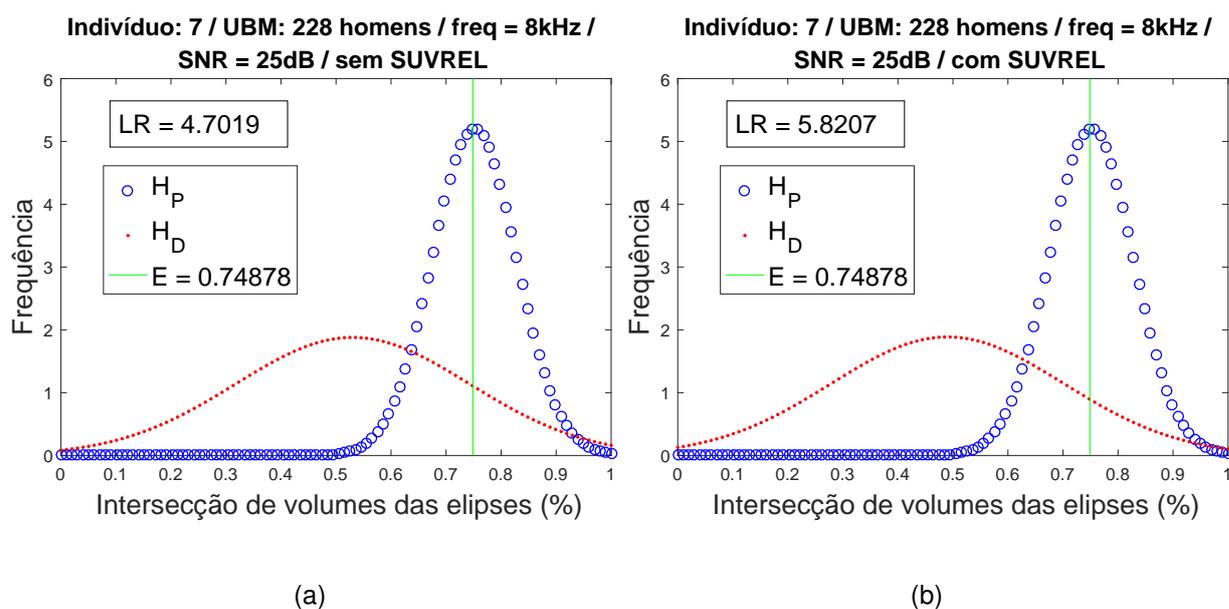
Fonte: produção própria.

Figura A.26. – Cálculo de LR no cenário apresentado no título de cada gráfico.



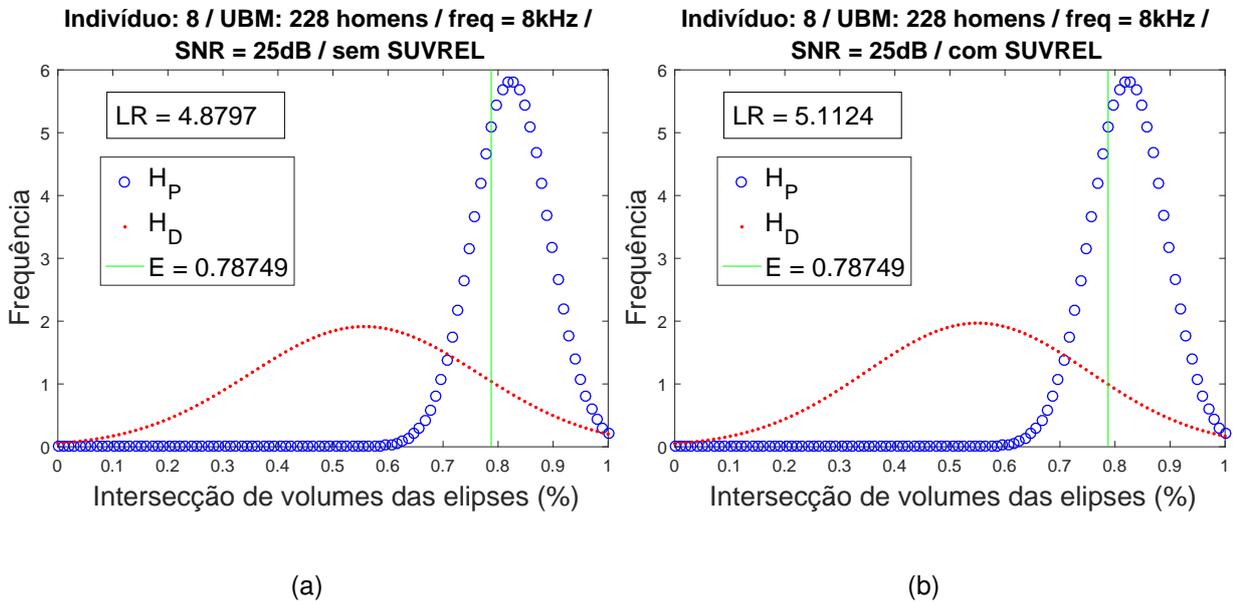
Fonte: produção própria.

Figura A.27. – Cálculo de LR no cenário apresentado no título de cada gráfico.



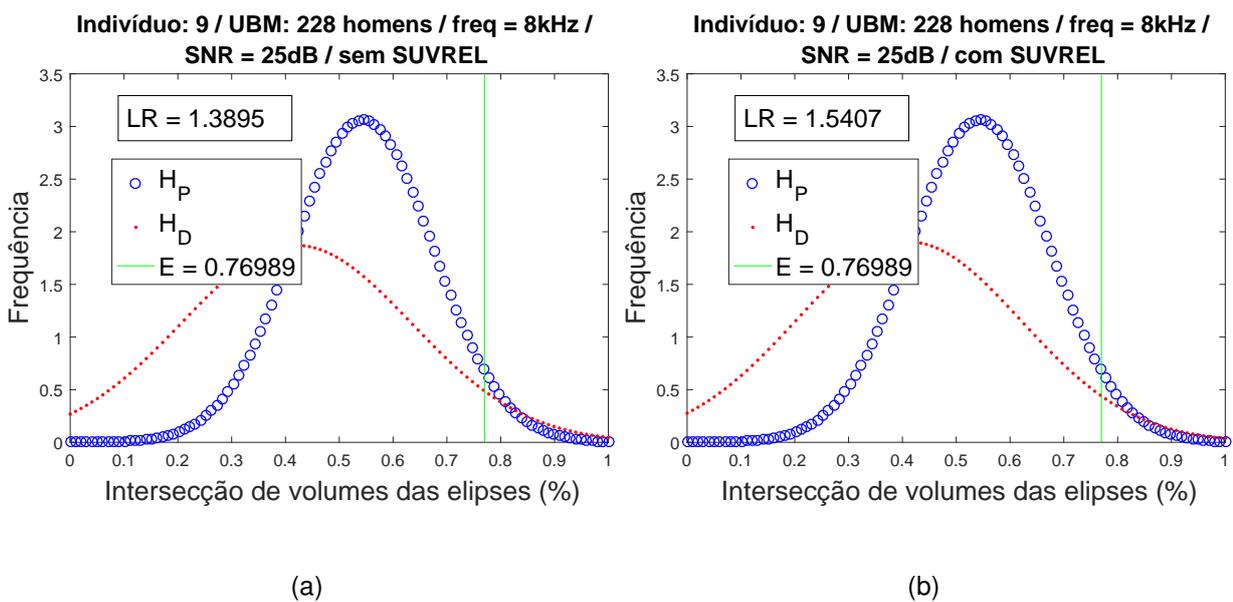
Fonte: produção própria.

Figura A.28. – Cálculo de LR no cenário apresentado no título de cada gráfico.



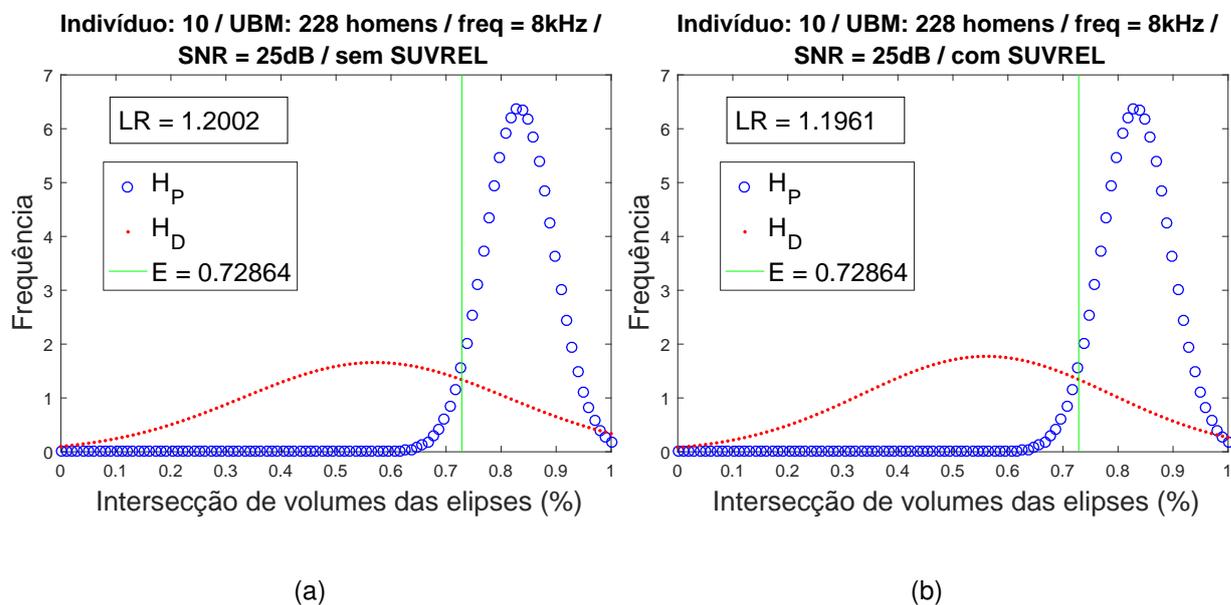
Fonte: produção própria.

Figura A.29. – Cálculo de LR no cenário apresentado no título de cada gráfico.



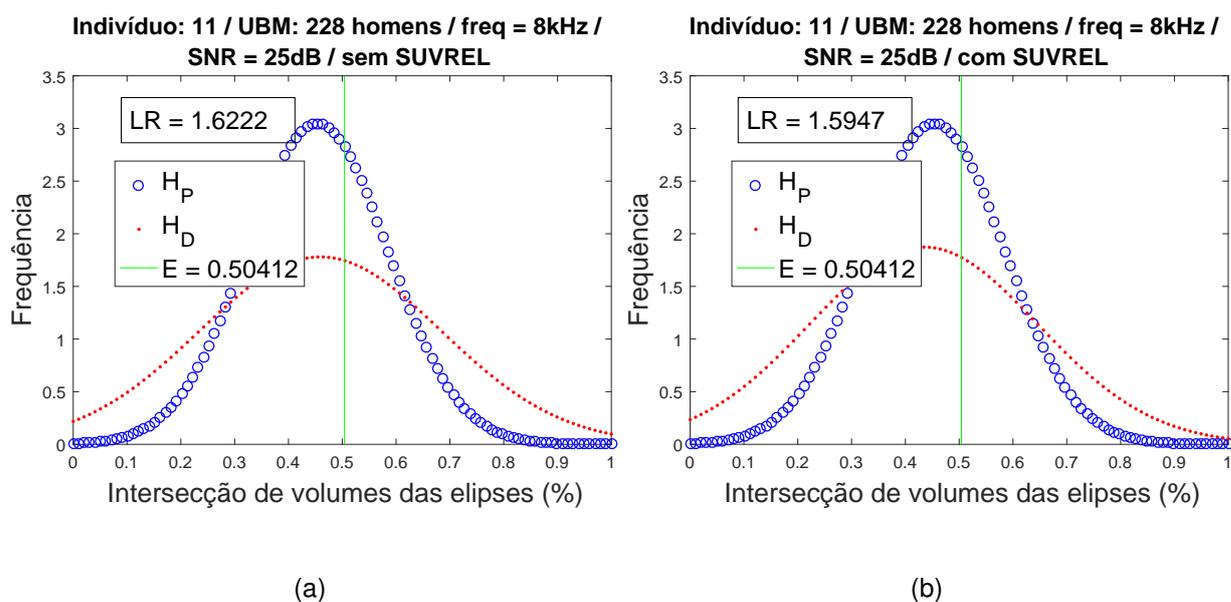
Fonte: produção própria.

Figura A.30. – Cálculo de LR no cenário apresentado no título de cada gráfico.



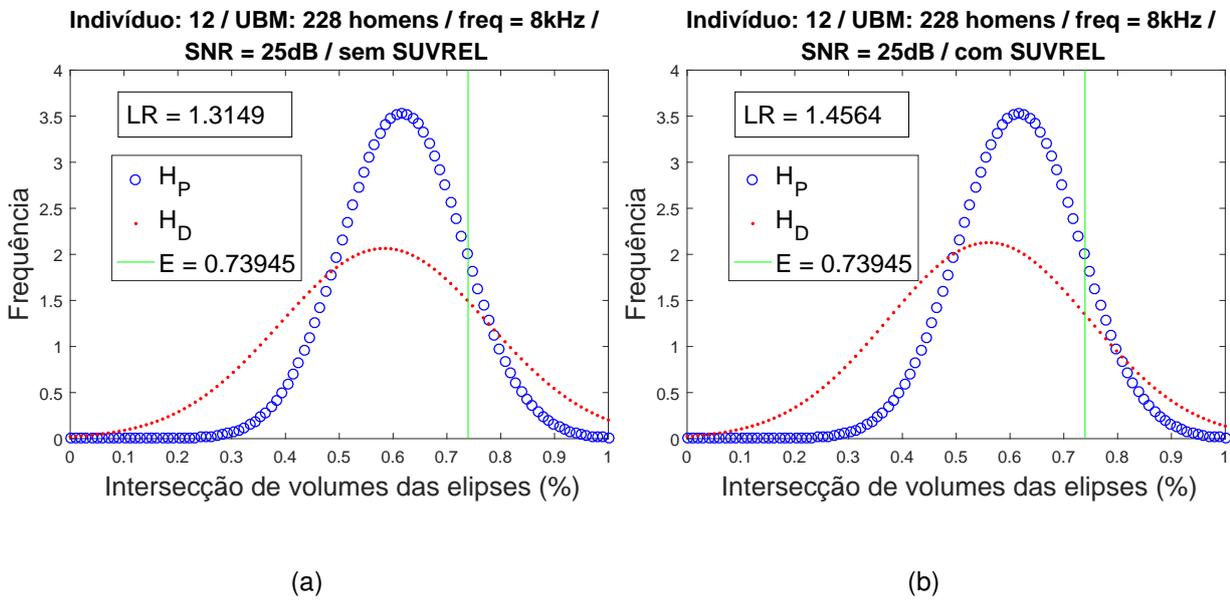
Fonte: produção própria.

Figura A.31. – Cálculo de LR no cenário apresentado no título de cada gráfico.



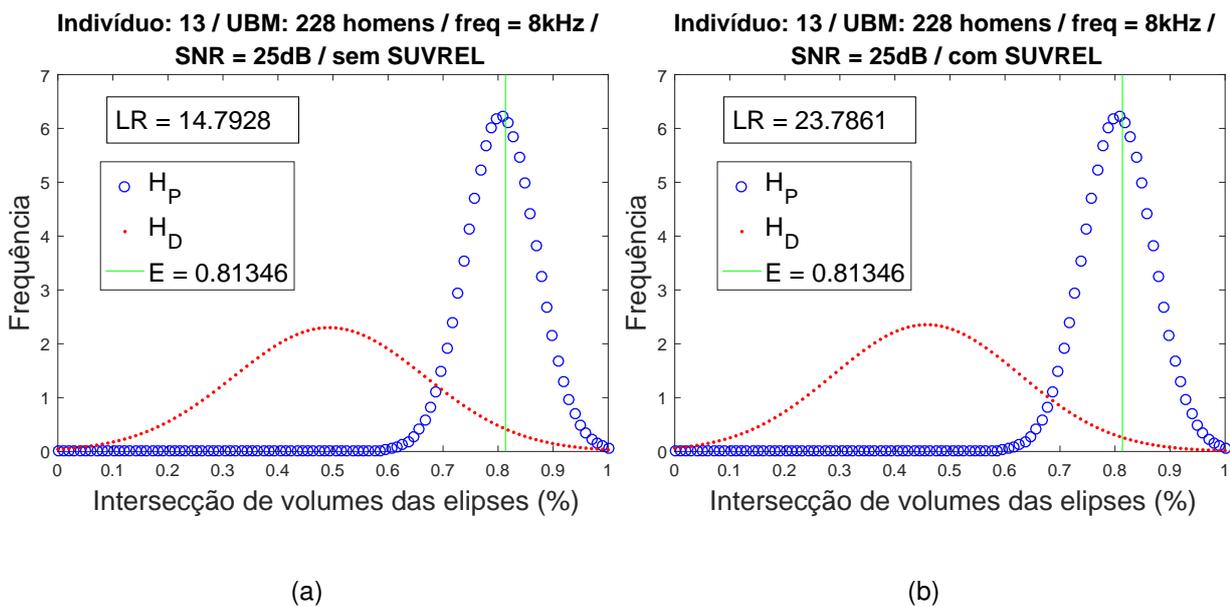
Fonte: produção própria.

Figura A.32. – Cálculo de LR no cenário apresentado no título de cada gráfico.



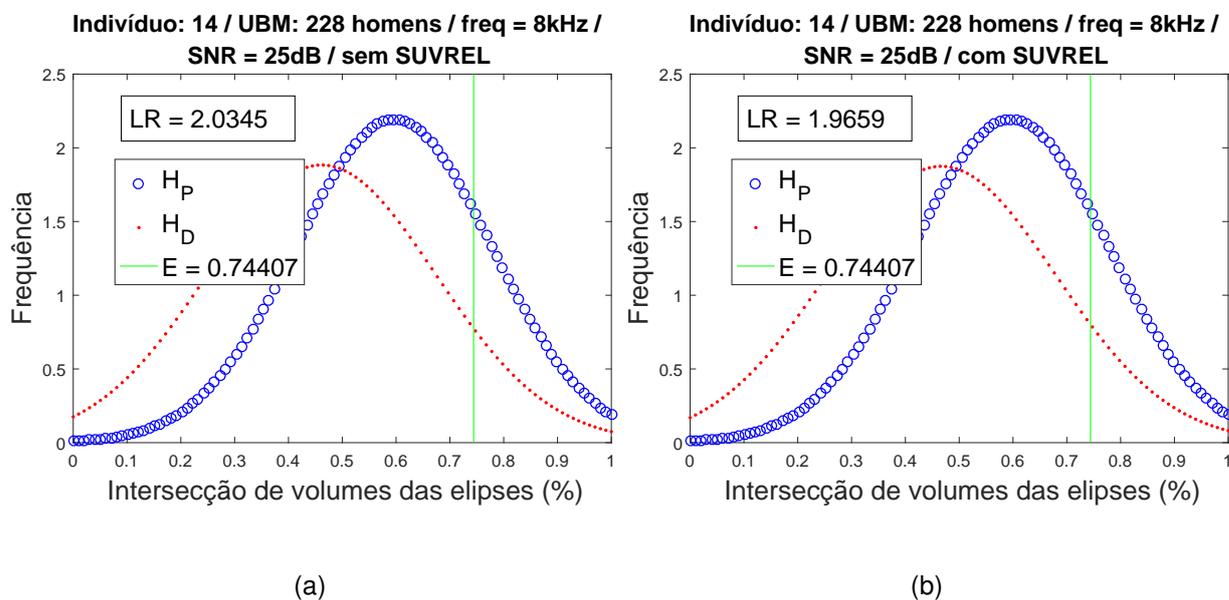
Fonte: produção própria.

Figura A.33. – Cálculo de LR no cenário apresentado no título de cada gráfico.



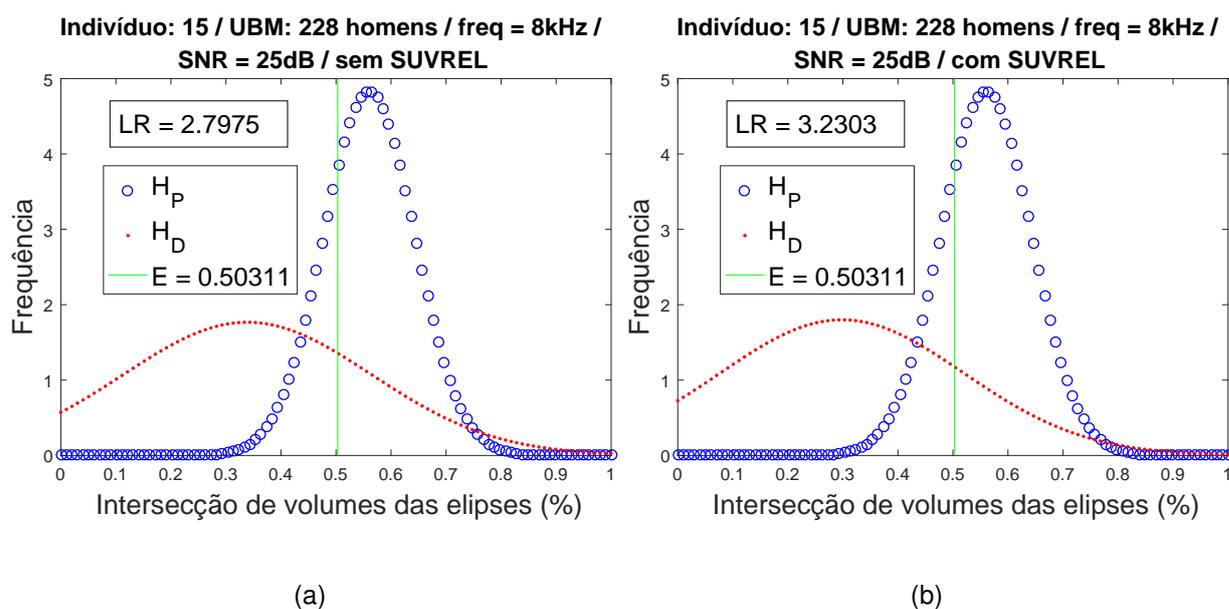
Fonte: produção própria.

Figura A.34. – Cálculo de LR no cenário apresentado no título de cada gráfico.



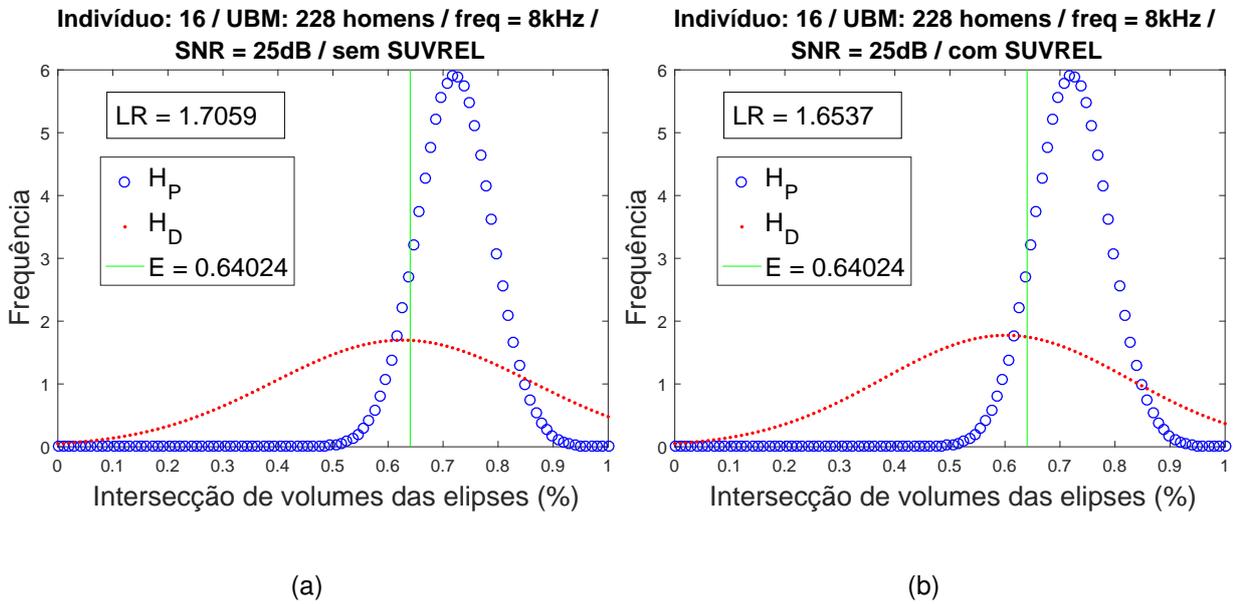
Fonte: produção própria.

Figura A.35. – Cálculo de LR no cenário apresentado no título de cada gráfico.



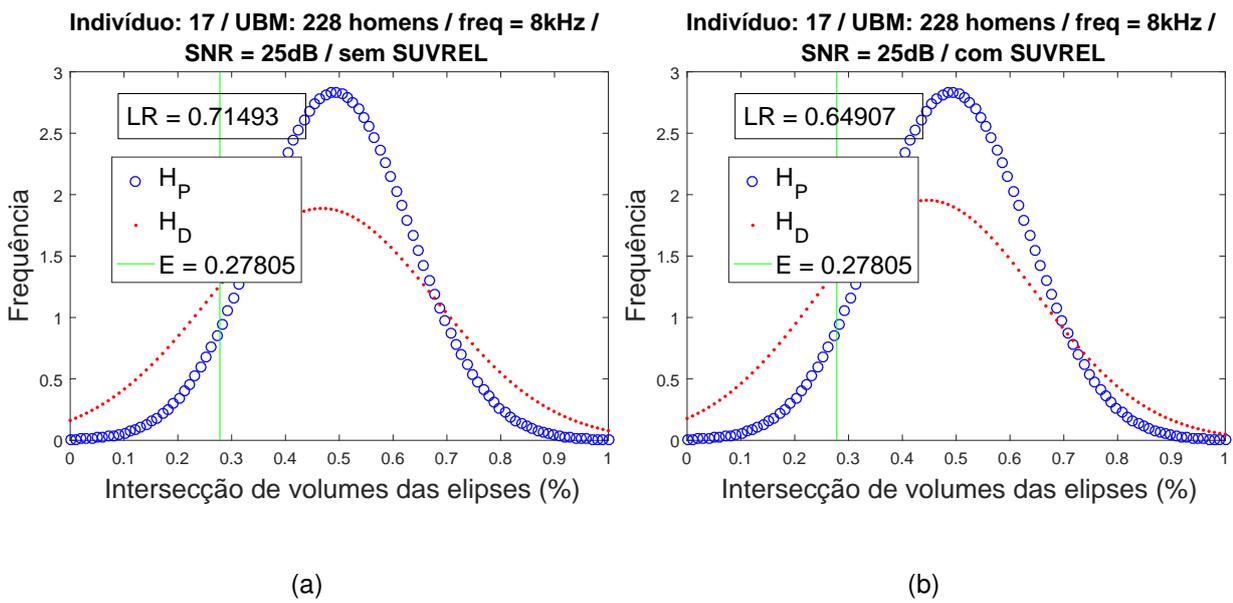
Fonte: produção própria.

Figura A.36. – Cálculo de LR no cenário apresentado no título de cada gráfico.



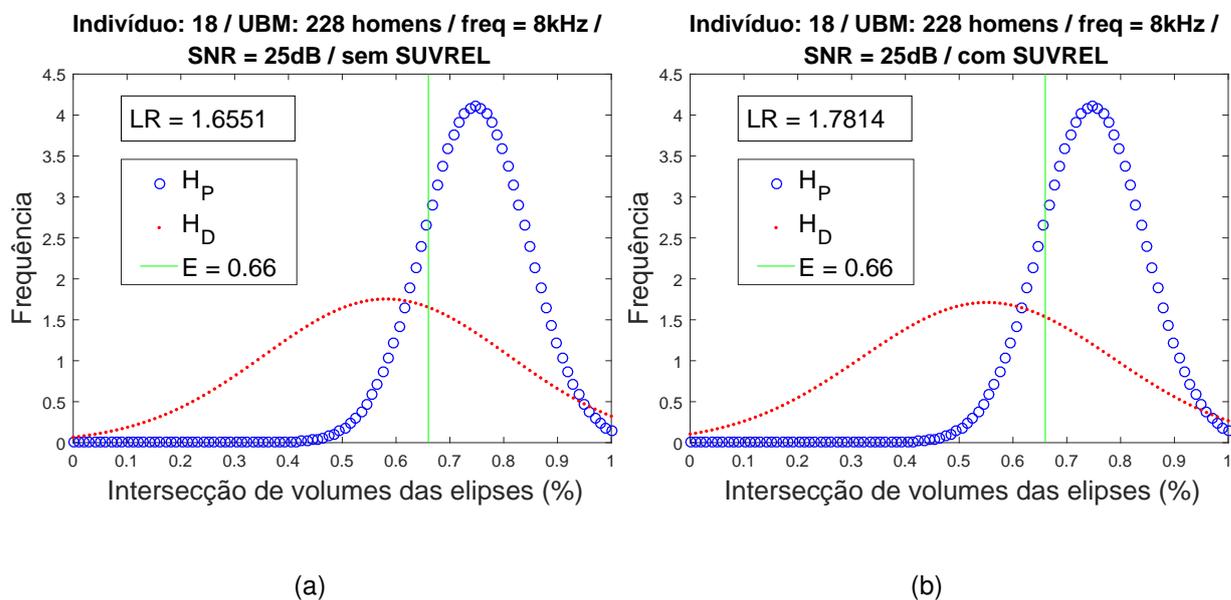
Fonte: produção própria.

Figura A.37. – Cálculo de LR no cenário apresentado no título de cada gráfico.



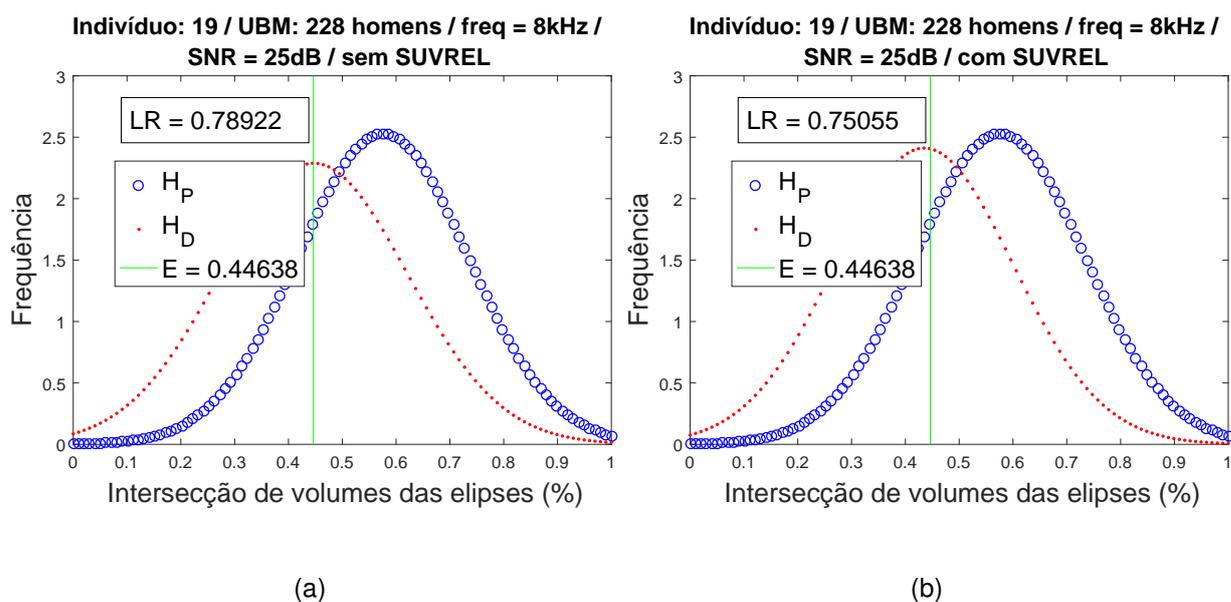
Fonte: produção própria.

Figura A.38. – Cálculo de LR no cenário apresentado no título de cada gráfico.



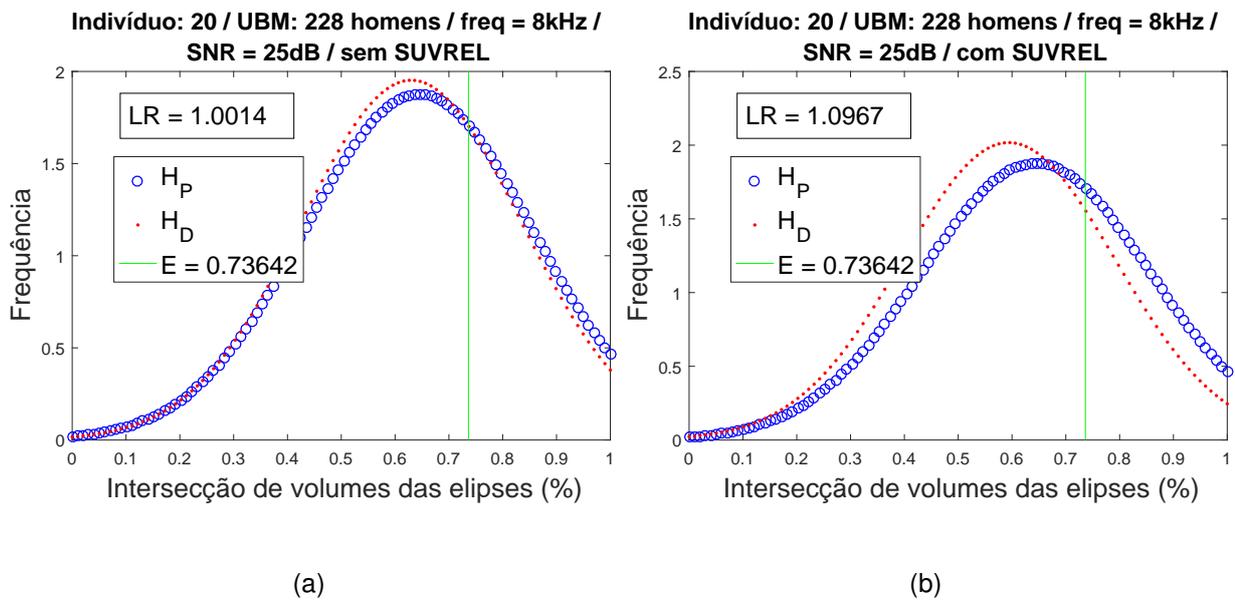
Fonte: produção própria.

Figura A.39. – Cálculo de LR no cenário apresentado no título de cada gráfico.



Fonte: produção própria.

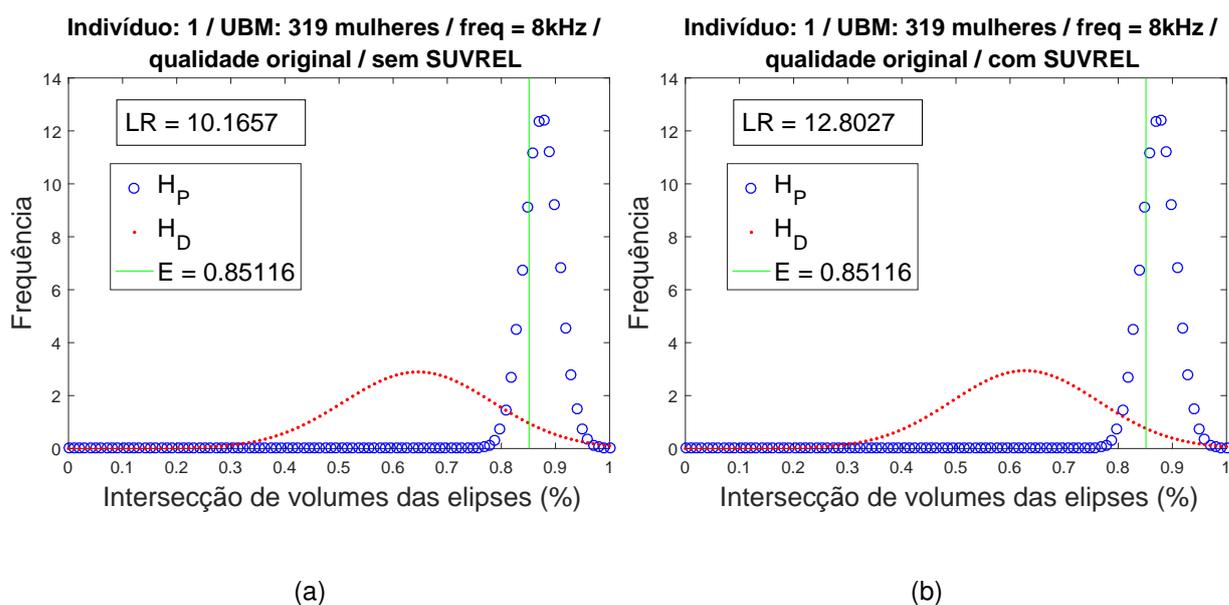
Figura A.40. – Cálculo de LR no cenário apresentado no título de cada gráfico.



Fonte: produção própria.

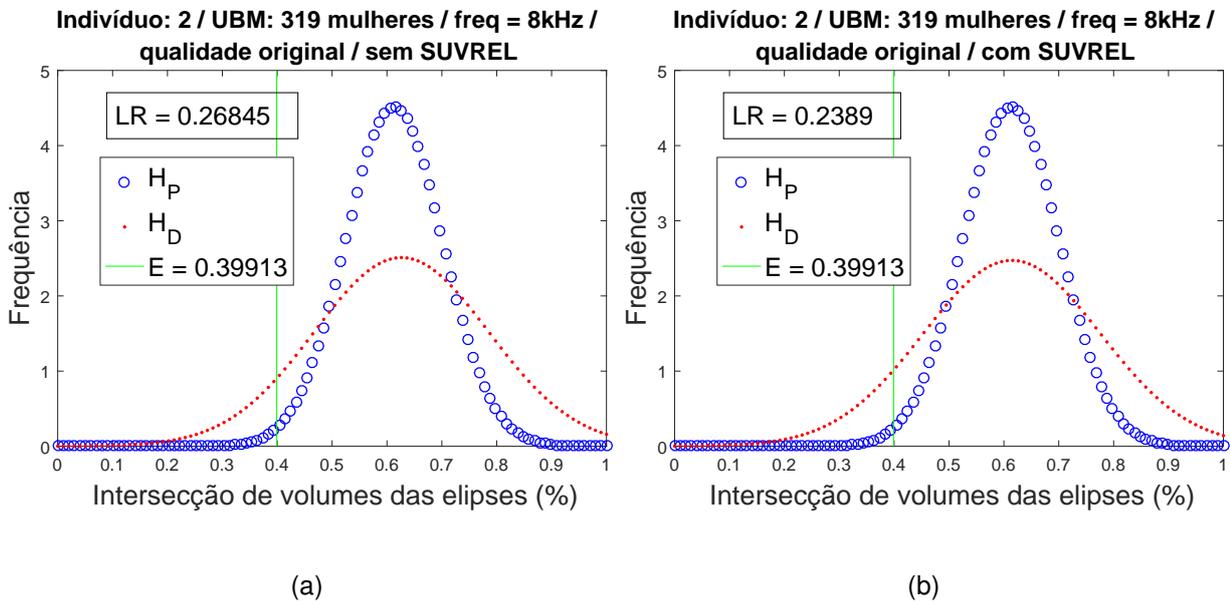
A.3. UBM: 319 mulheres / frequência: 8kHz / qualidade original

Figura A.41. – Cálculo de LR no cenário apresentado no título de cada gráfico.



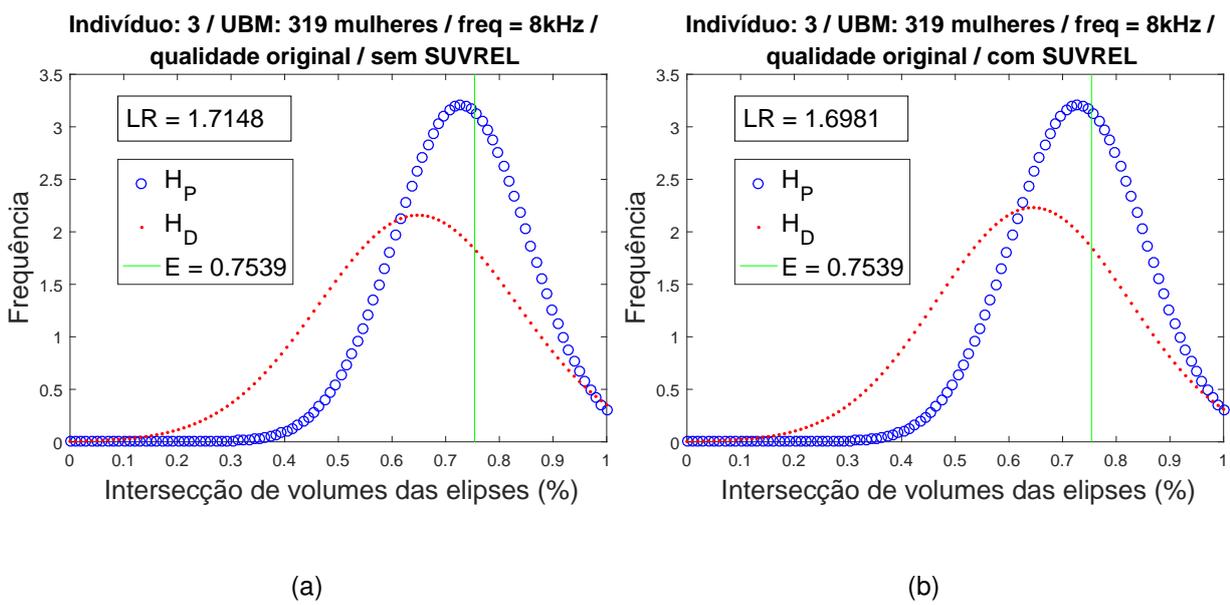
Fonte: produção própria.

Figura A.42. – Cálculo de LR no cenário apresentado no título de cada gráfico.



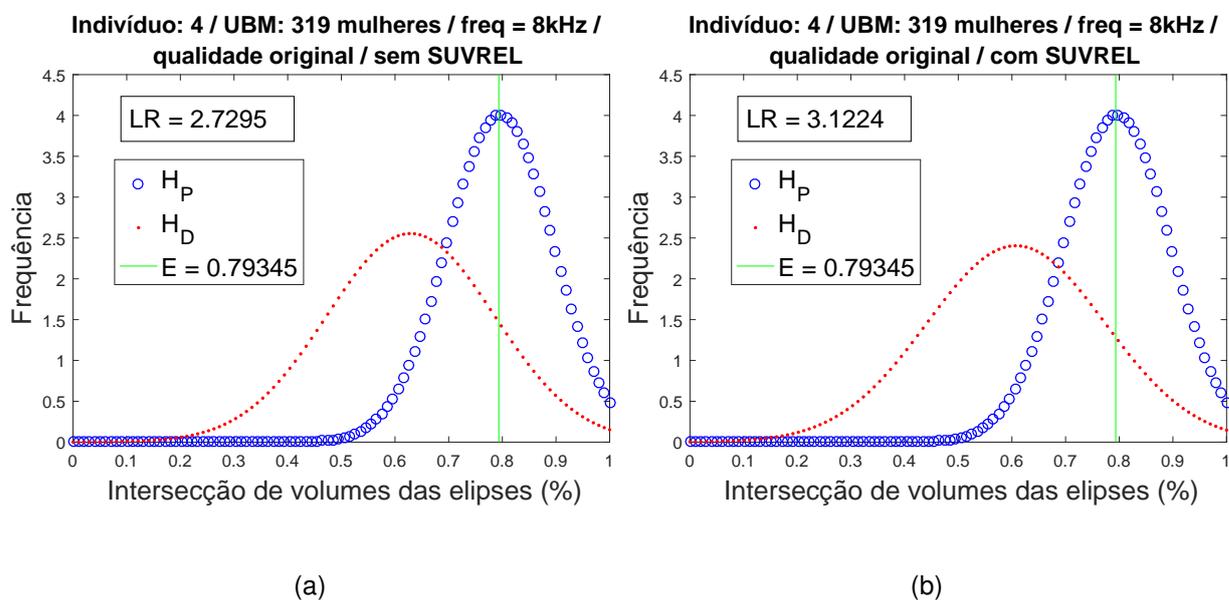
Fonte: produção própria.

Figura A.43. – Cálculo de LR no cenário apresentado no título de cada gráfico.



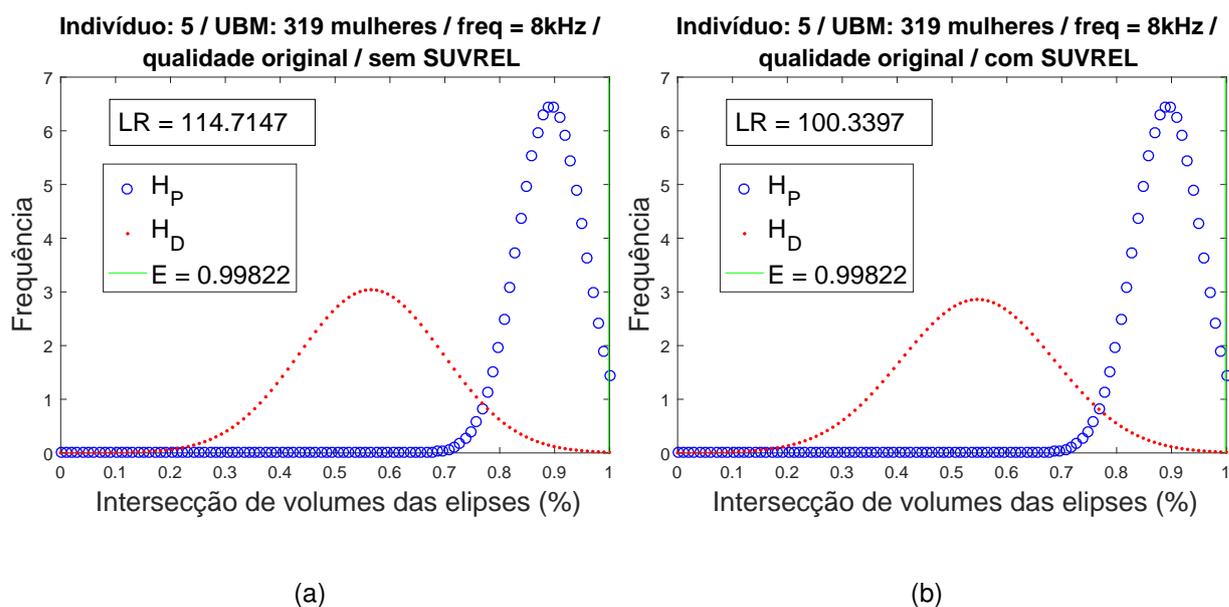
Fonte: produção própria.

Figura A.44. – Cálculo de LR no cenário apresentado no título de cada gráfico.



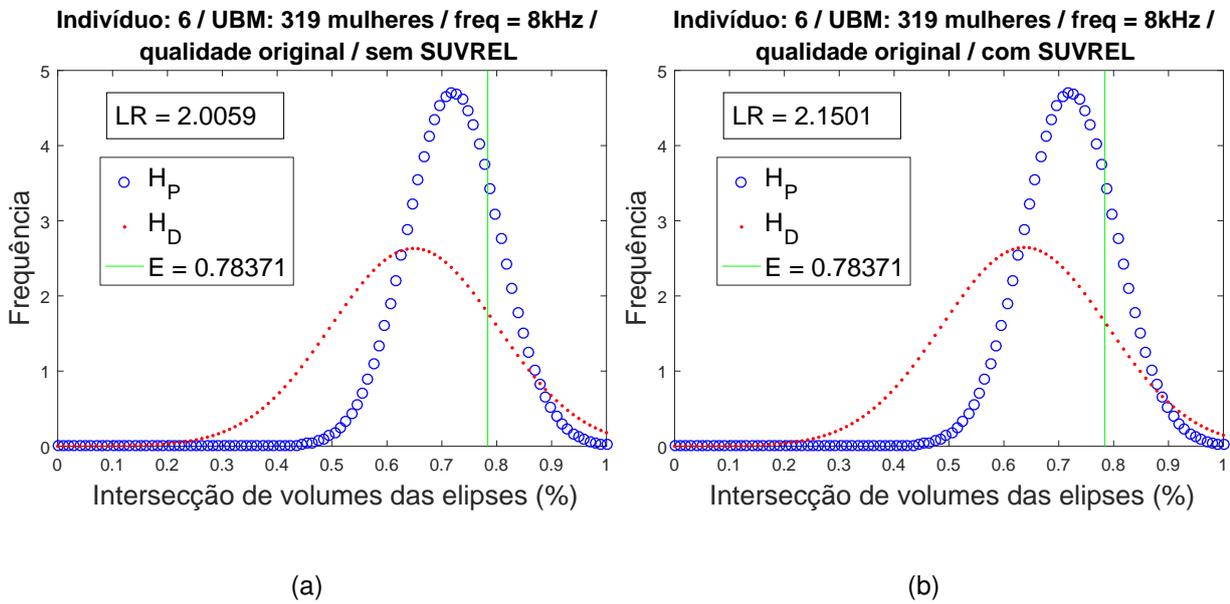
Fonte: produção própria.

Figura A.45. – Cálculo de LR no cenário apresentado no título de cada gráfico.



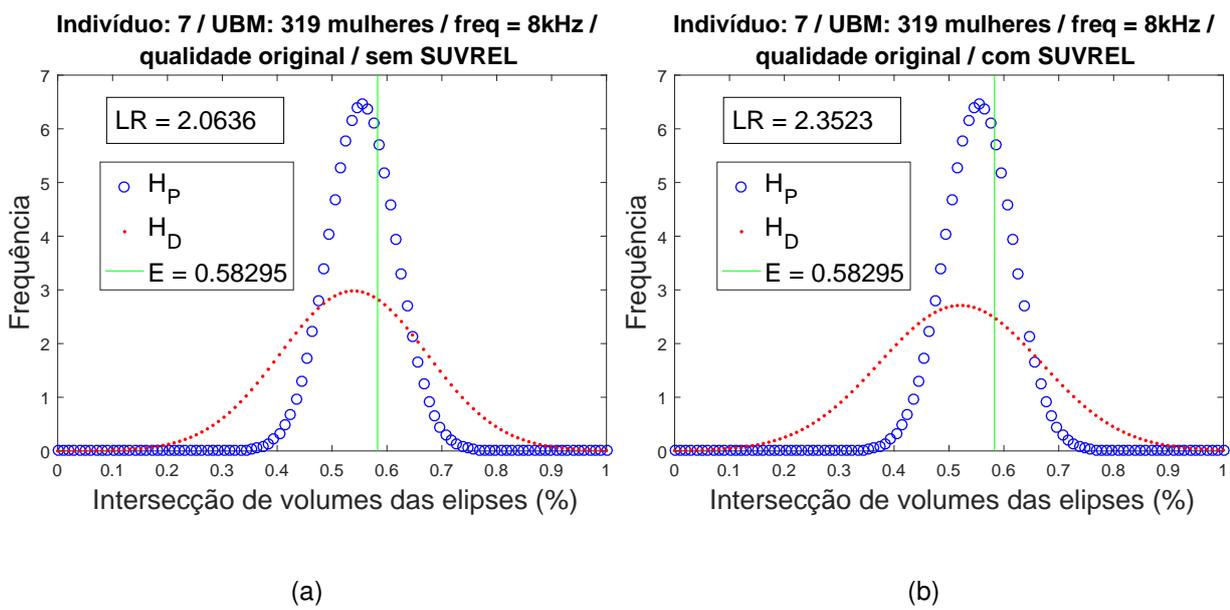
Fonte: produção própria.

Figura A.46. – Cálculo de LR no cenário apresentado no título de cada gráfico.



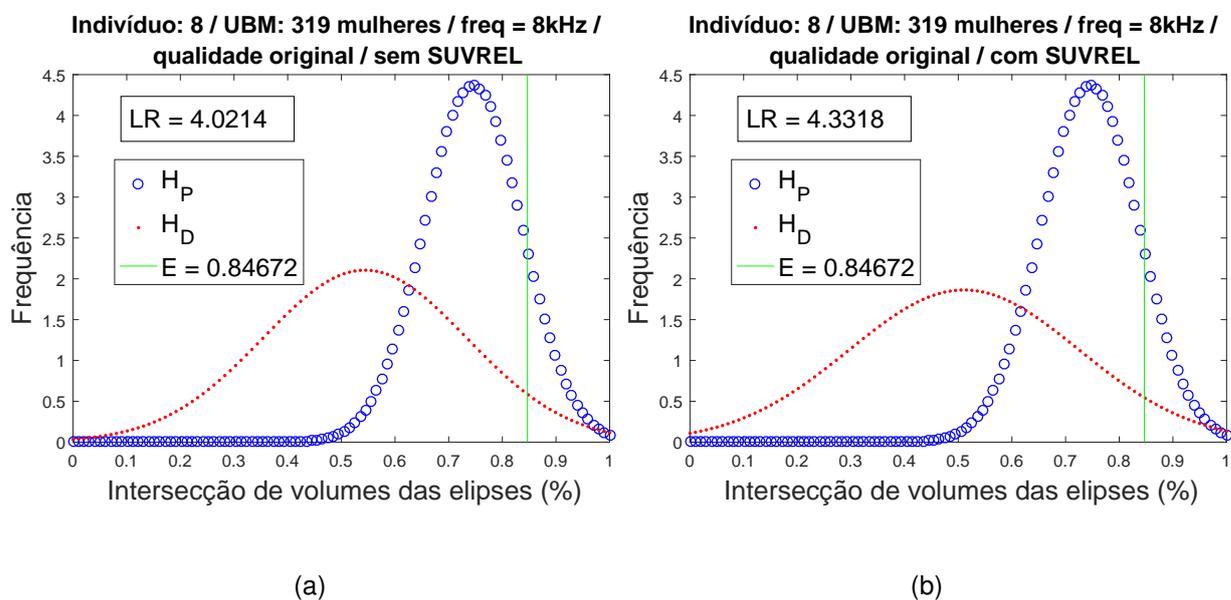
Fonte: produção própria.

Figura A.47. – Cálculo de LR no cenário apresentado no título de cada gráfico.



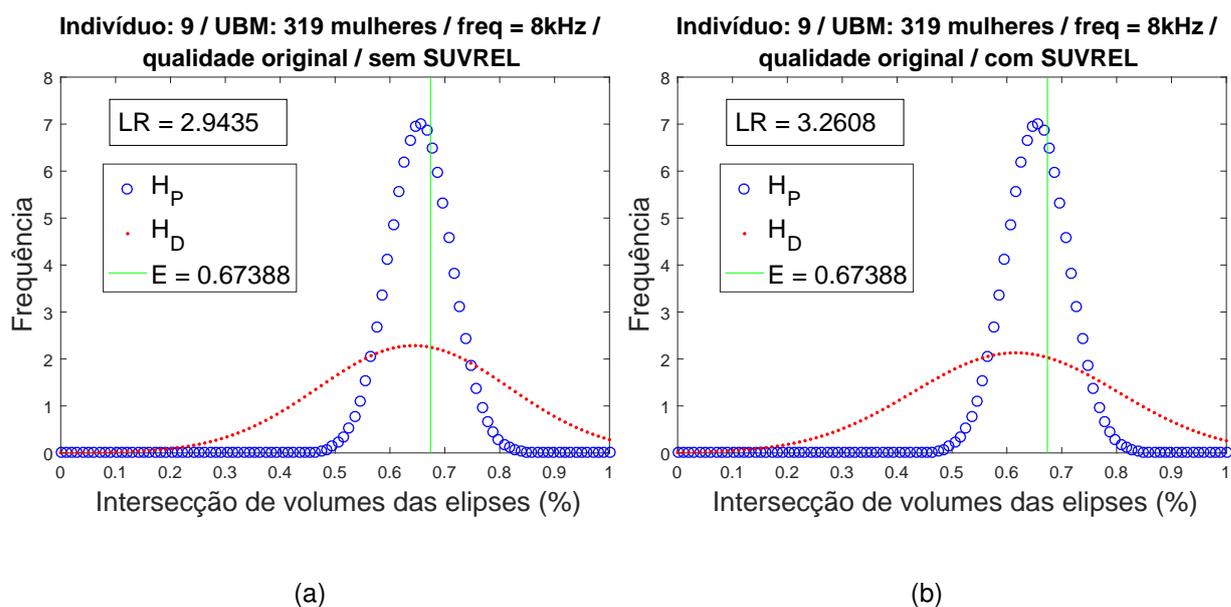
Fonte: produção própria.

Figura A.48. – Cálculo de LR no cenário apresentado no título de cada gráfico.



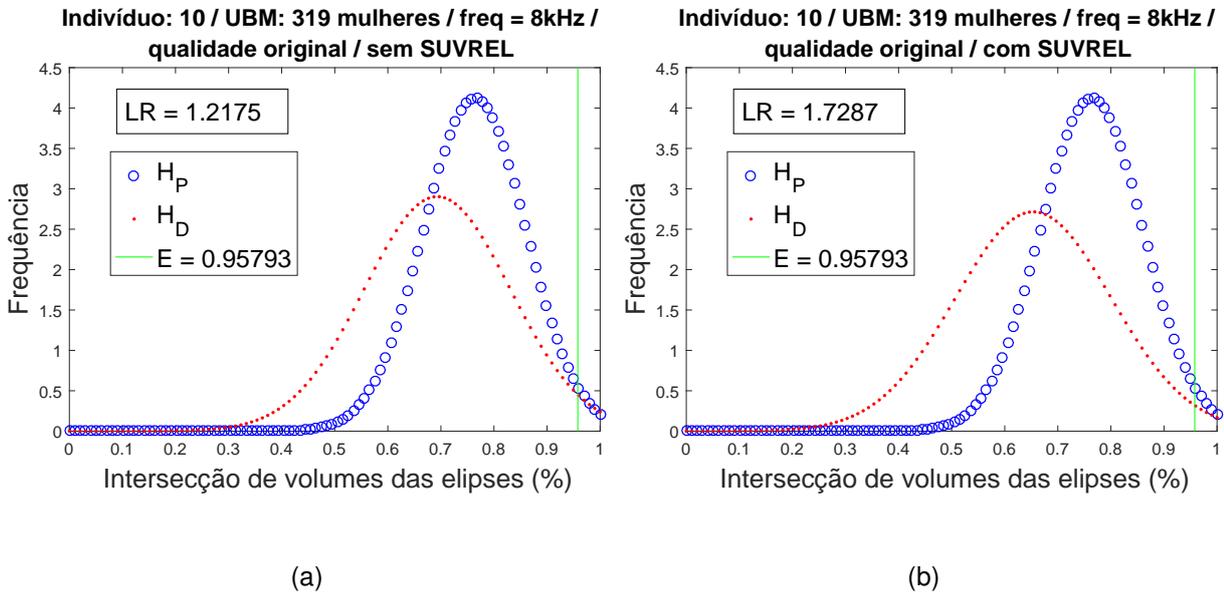
Fonte: produção própria.

Figura A.49. – Cálculo de LR no cenário apresentado no título de cada gráfico.



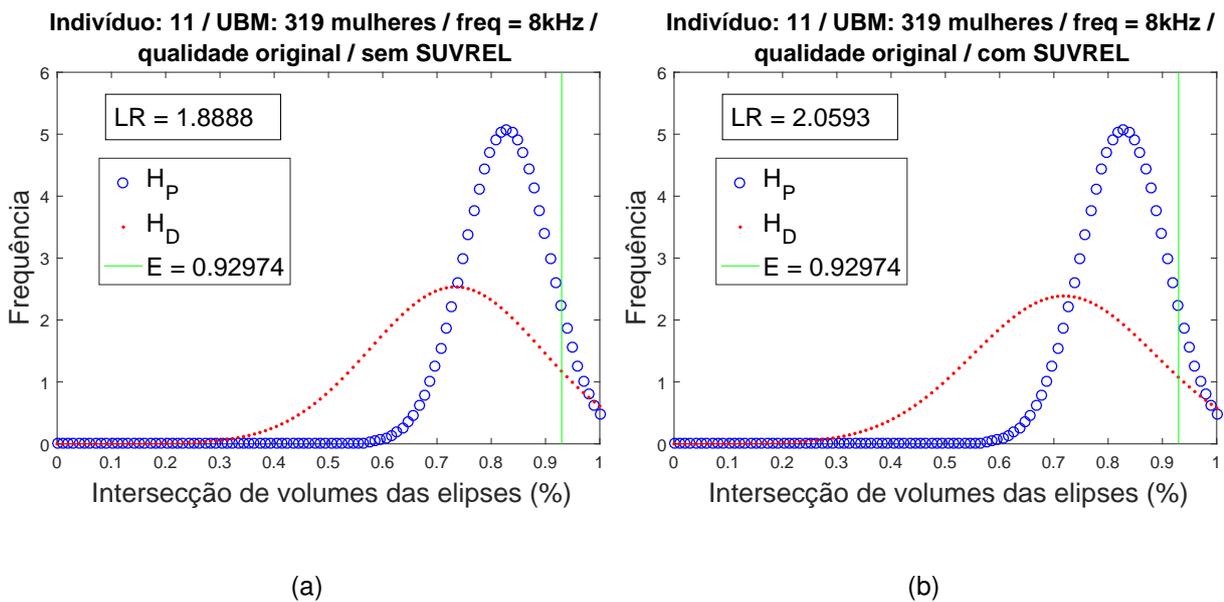
Fonte: produção própria.

Figura A.50. – Cálculo de LR no cenário apresentado no título de cada gráfico.



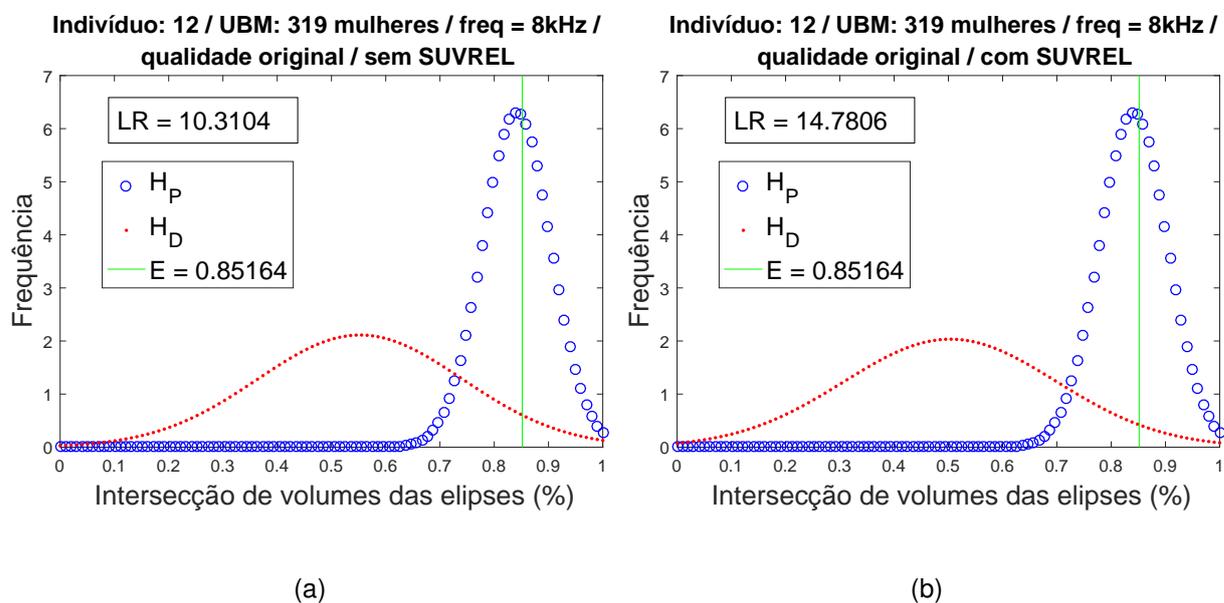
Fonte: produção própria.

Figura A.51. – Cálculo de LR no cenário apresentado no título de cada gráfico.



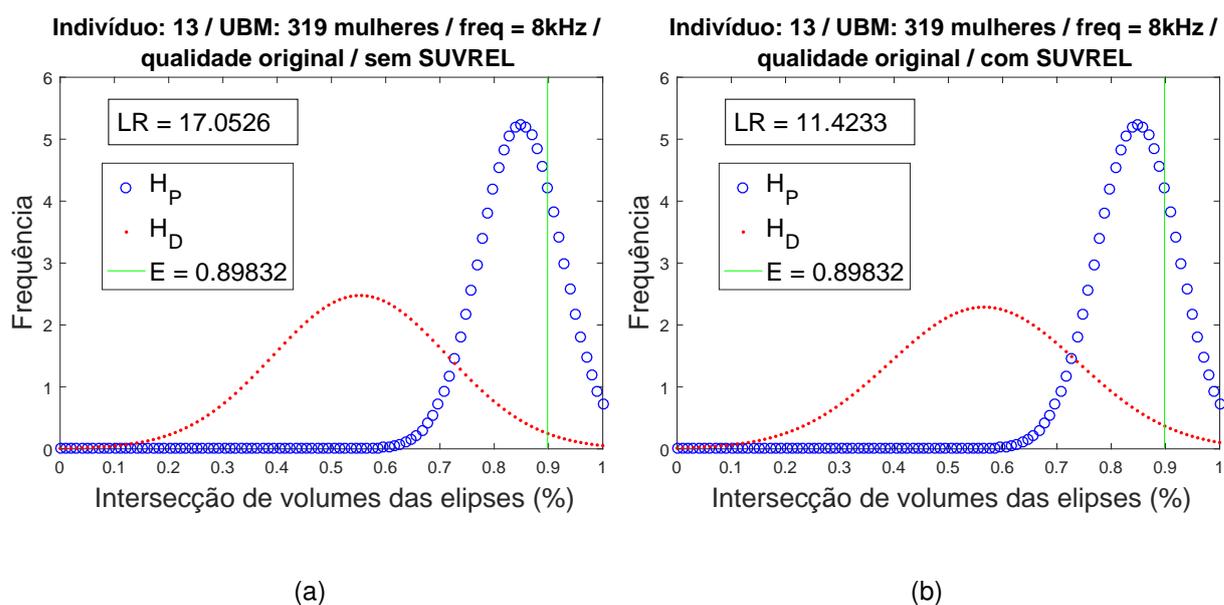
Fonte: produção própria.

Figura A.52. – Cálculo de LR no cenário apresentado no título de cada gráfico.



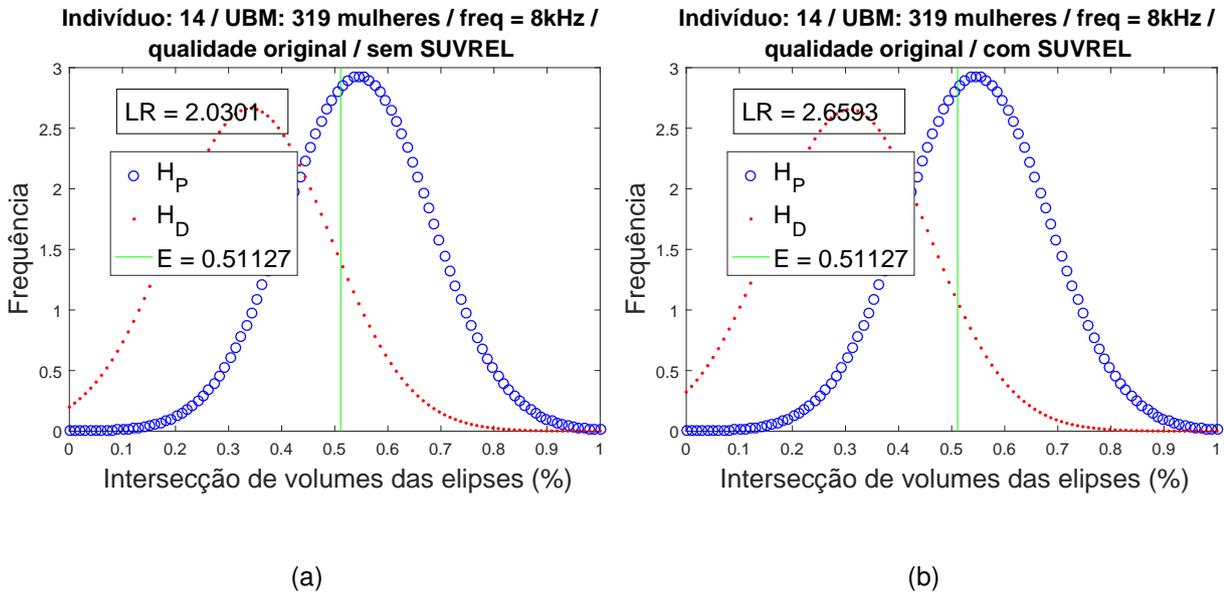
Fonte: produção própria.

Figura A.53. – Cálculo de LR no cenário apresentado no título de cada gráfico.



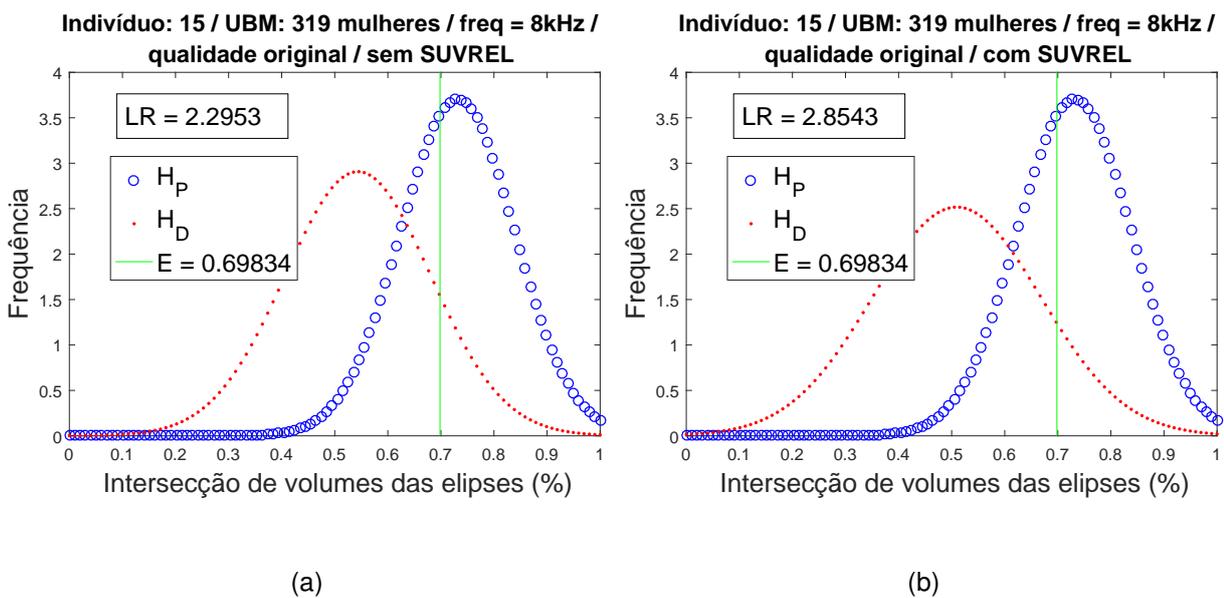
Fonte: produção própria.

Figura A.54. – Cálculo de LR no cenário apresentado no título de cada gráfico.



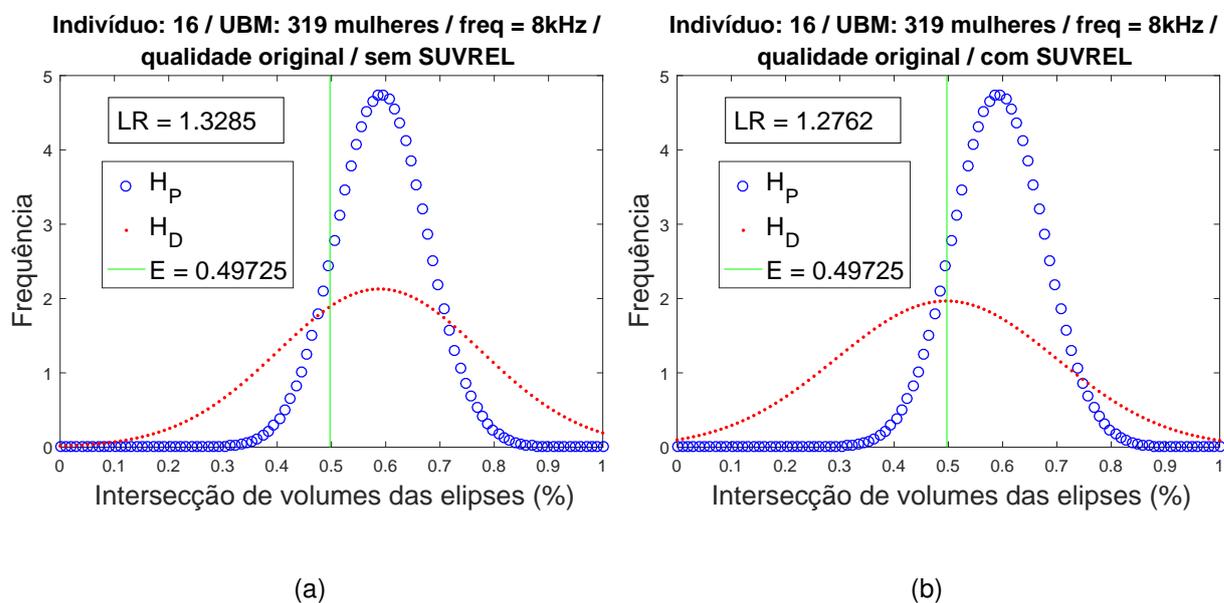
Fonte: produção própria.

Figura A.55. – Cálculo de LR no cenário apresentado no título de cada gráfico.



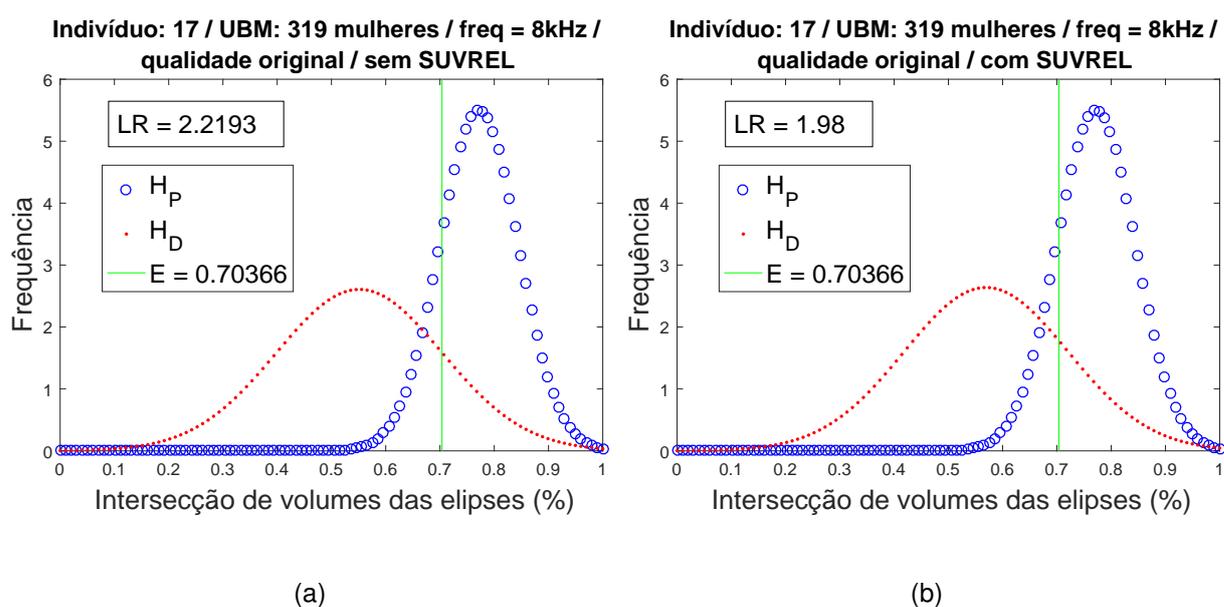
Fonte: produção própria.

Figura A.56. – Cálculo de LR no cenário apresentado no título de cada gráfico.



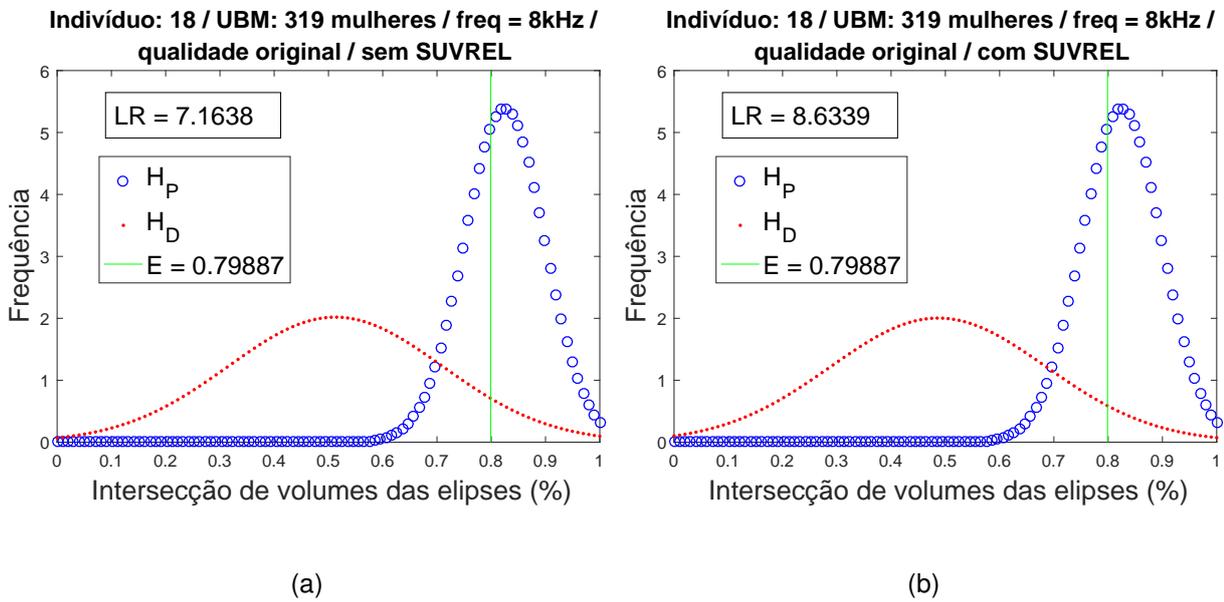
Fonte: produção própria.

Figura A.57. – Cálculo de LR no cenário apresentado no título de cada gráfico.



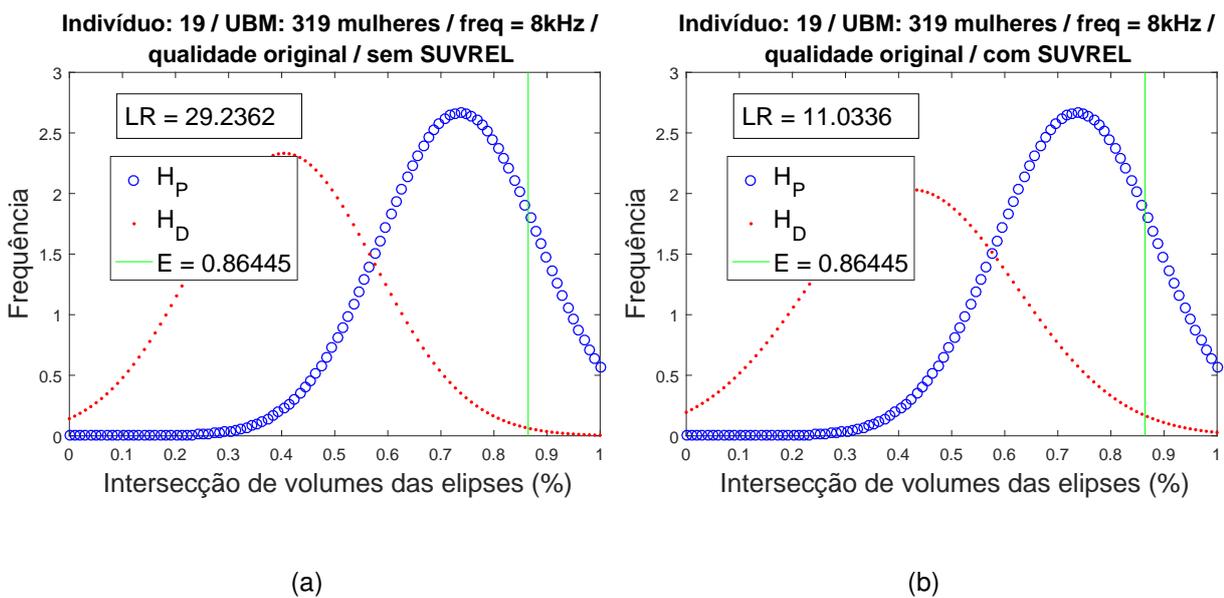
Fonte: produção própria.

Figura A.58. – Cálculo de LR no cenário apresentado no título de cada gráfico.



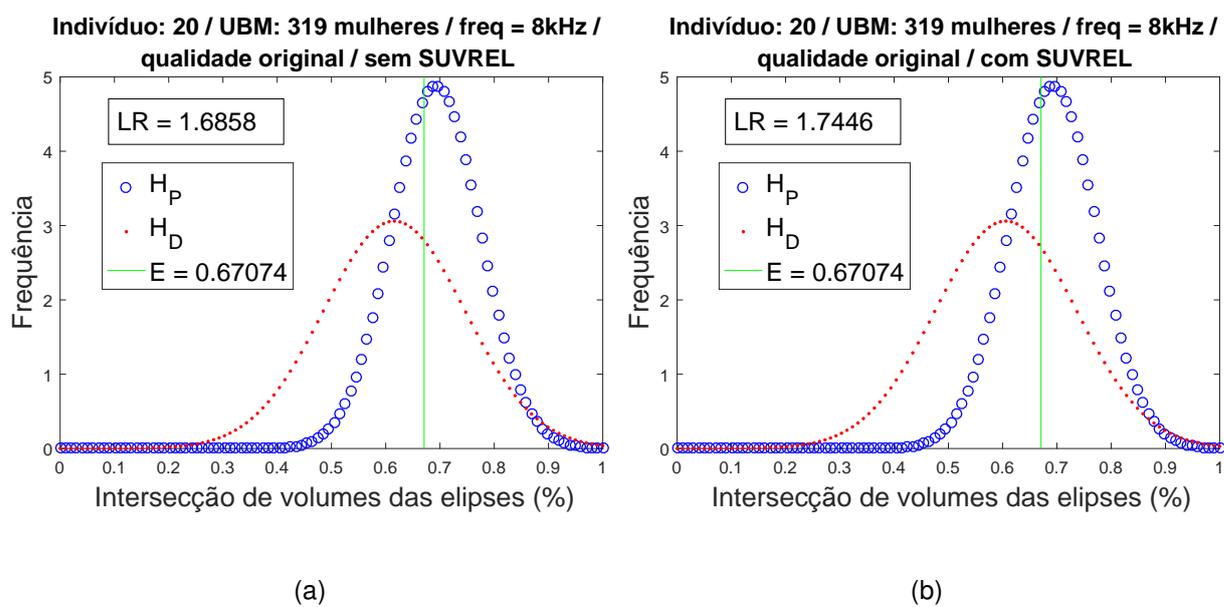
Fonte: produção própria.

Figura A.59. – Cálculo de LR no cenário apresentado no título de cada gráfico.



Fonte: produção própria.

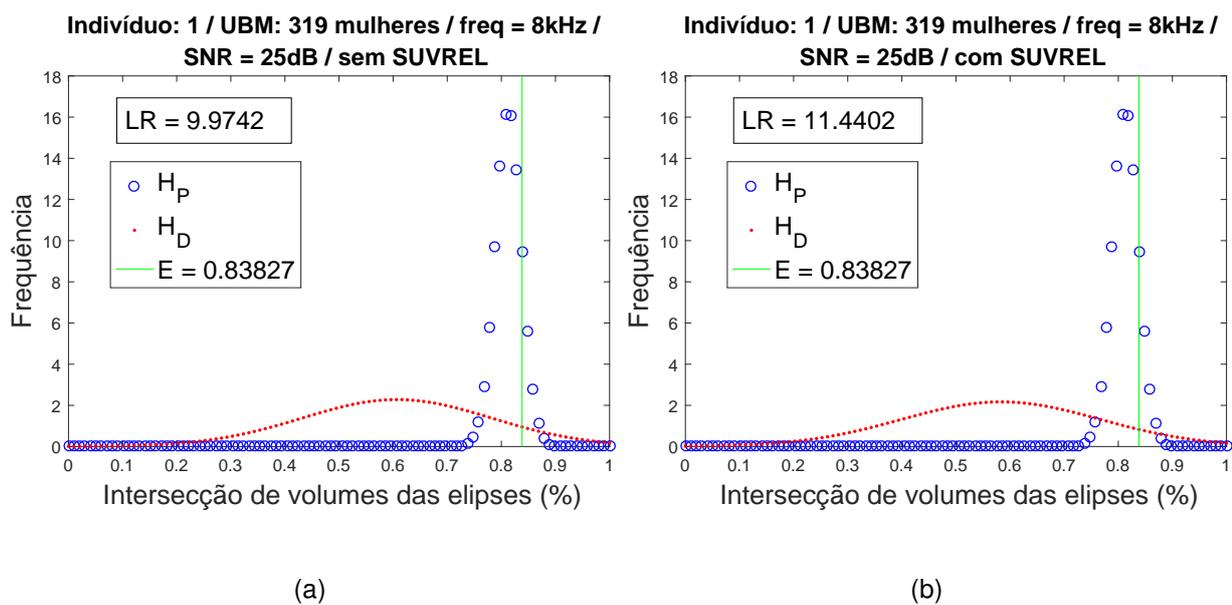
Figura A.60. – Cálculo de LR no cenário apresentado no título de cada gráfico.



Fonte: produção própria.

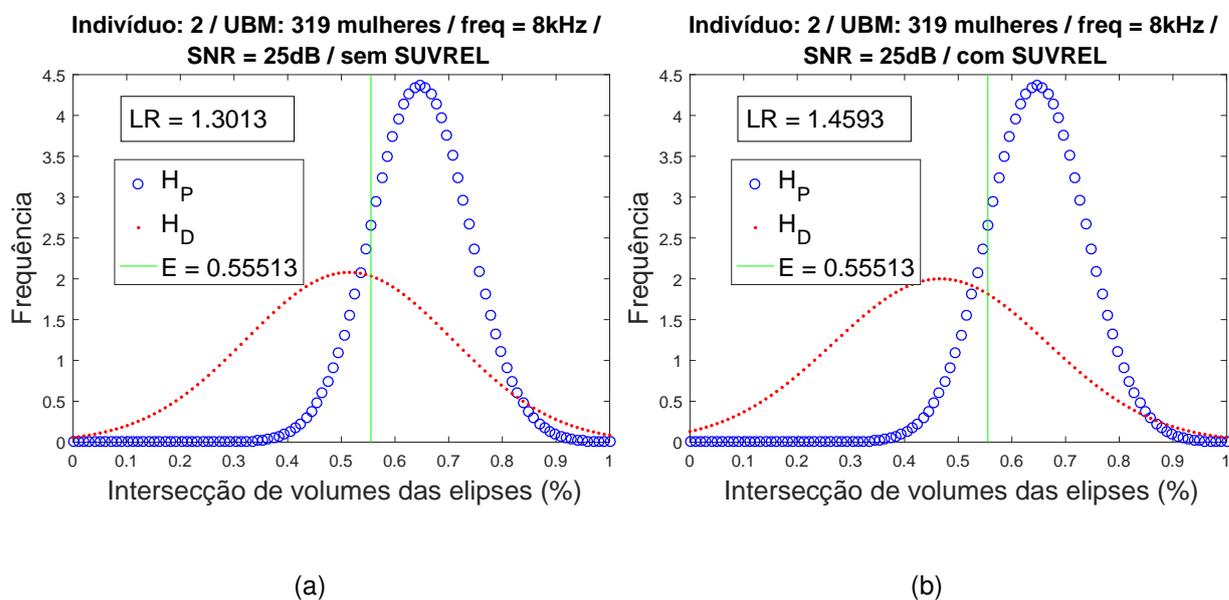
A.4. UBM: 319 mulheres / frequência: 8kHz / SNR: 25dB

Figura A.61. – Cálculo de LR no cenário apresentado no título de cada gráfico.



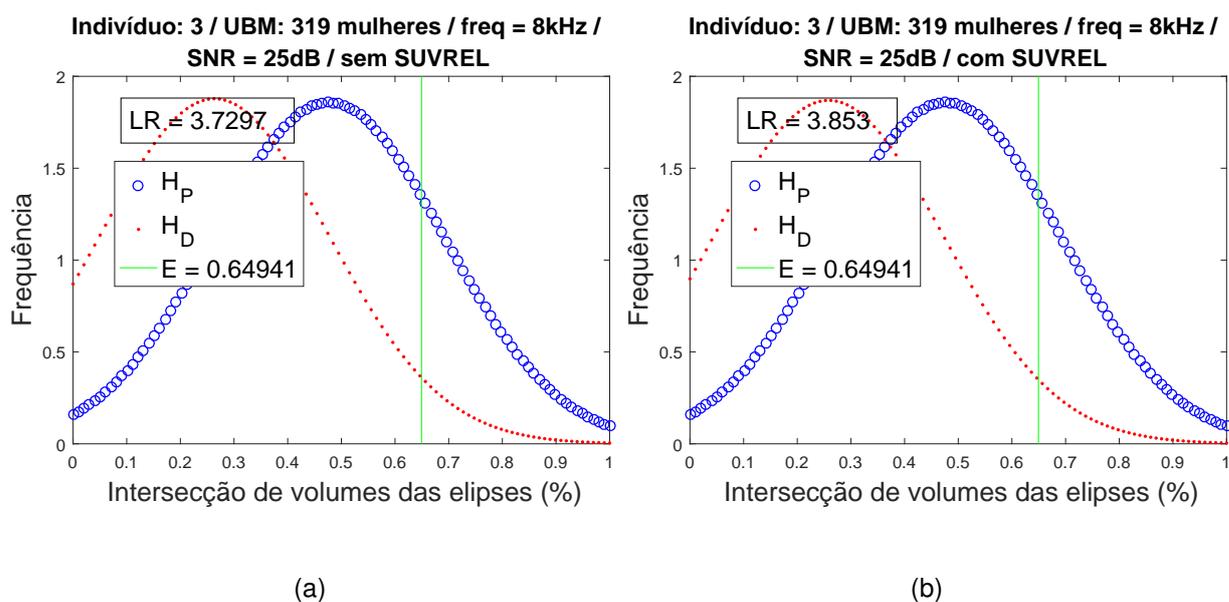
Fonte: produção própria.

Figura A.62. – Cálculo de LR no cenário apresentado no título de cada gráfico.



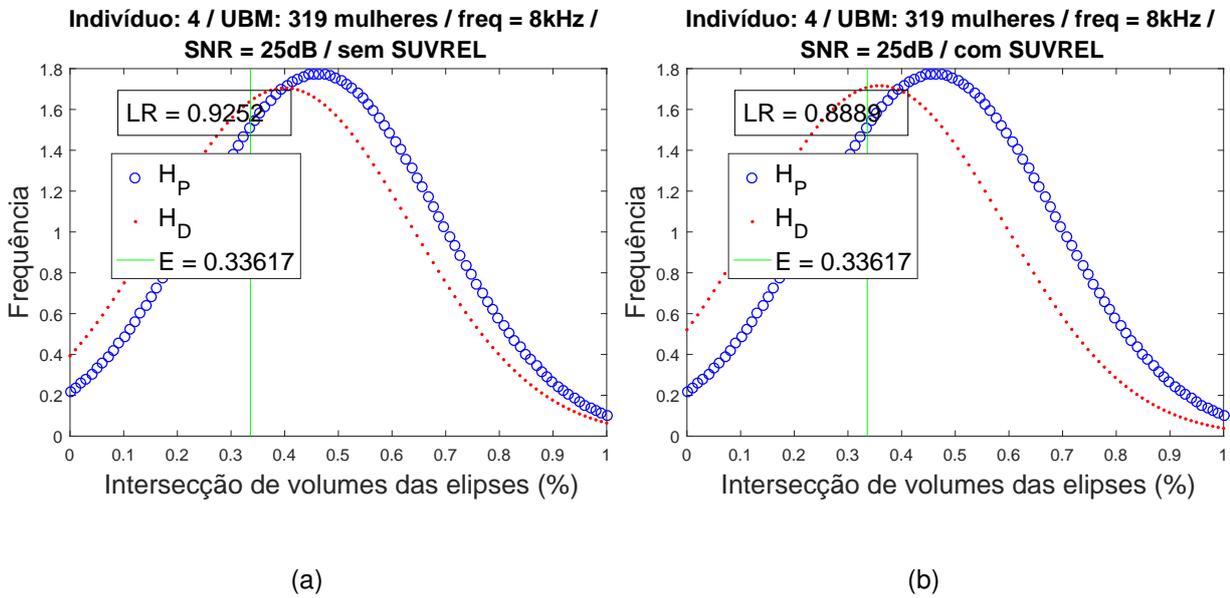
Fonte: produção própria.

Figura A.63. – Cálculo de LR no cenário apresentado no título de cada gráfico.



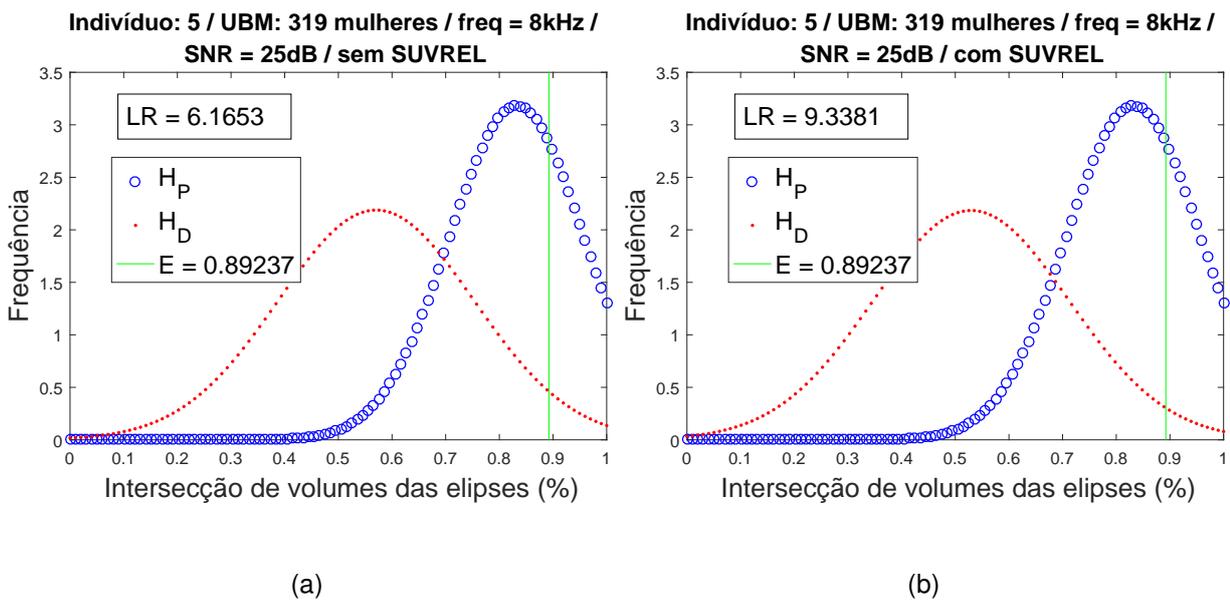
Fonte: produção própria.

Figura A.64. – Cálculo de LR no cenário apresentado no título de cada gráfico.



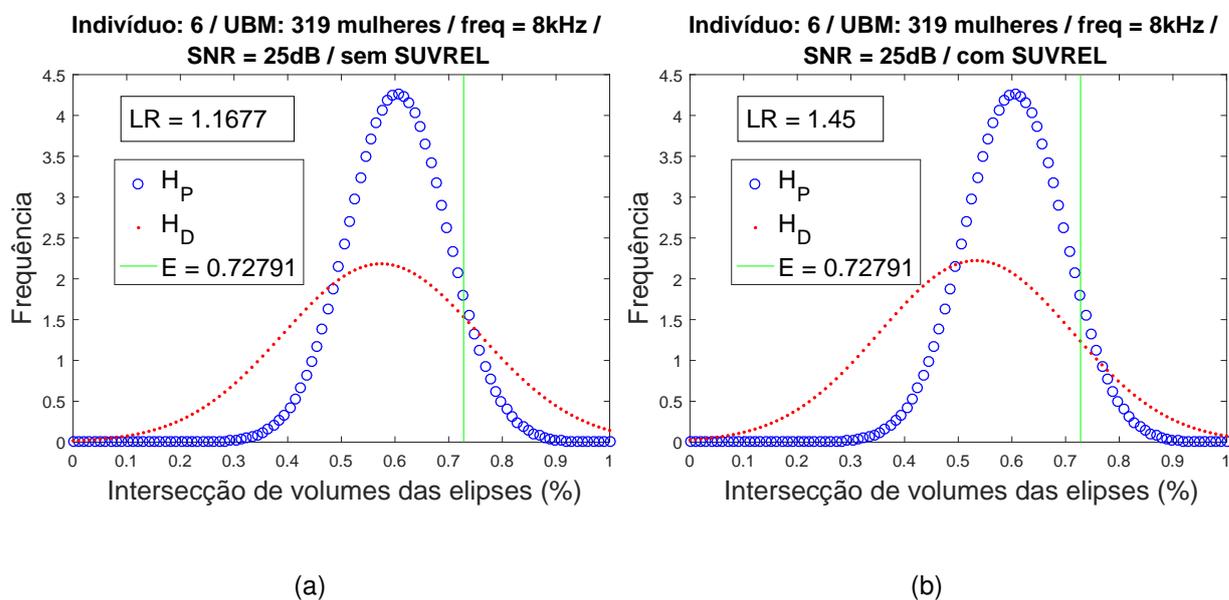
Fonte: produção própria.

Figura A.65. – Cálculo de LR no cenário apresentado no título de cada gráfico.



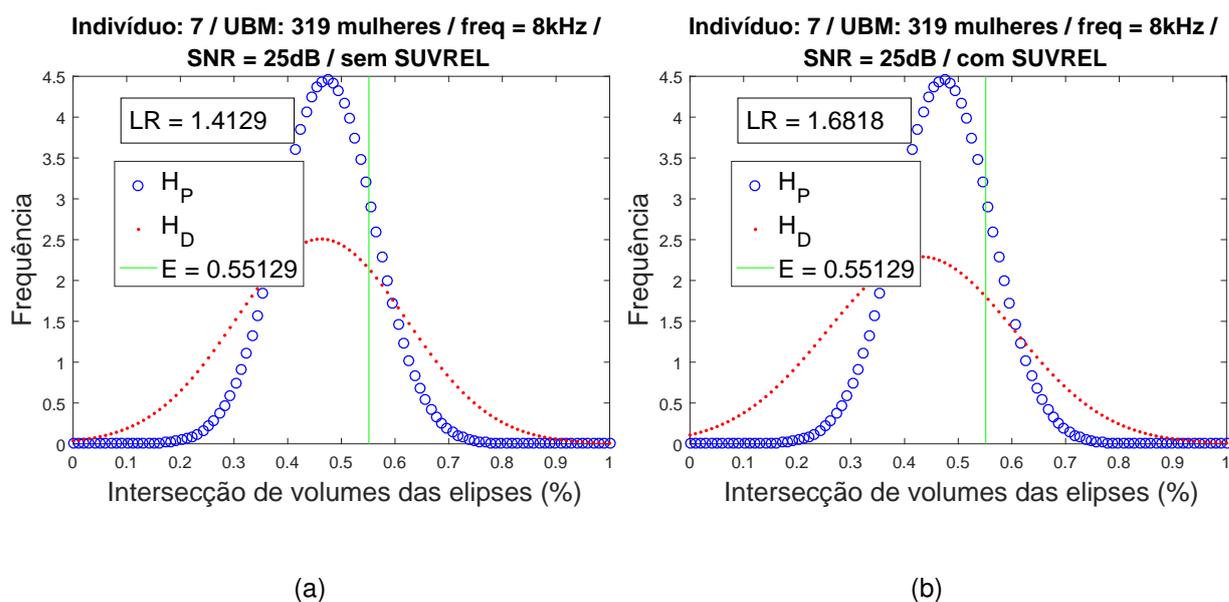
Fonte: produção própria.

Figura A.66. – Cálculo de LR no cenário apresentado no título de cada gráfico.



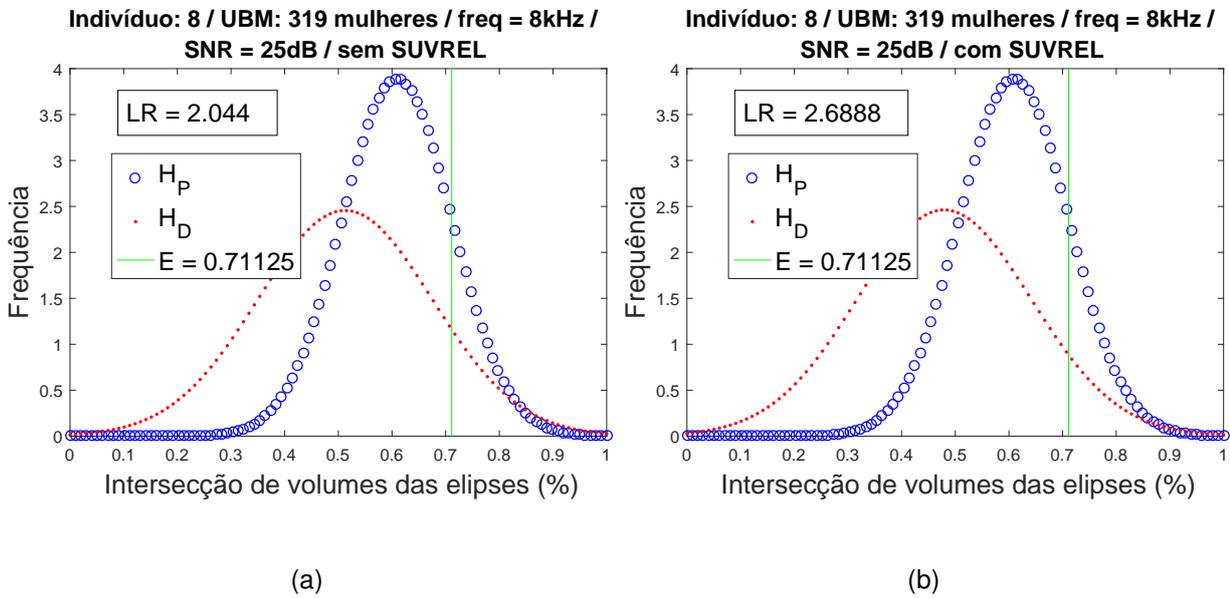
Fonte: produção própria.

Figura A.67. – Cálculo de LR no cenário apresentado no título de cada gráfico.



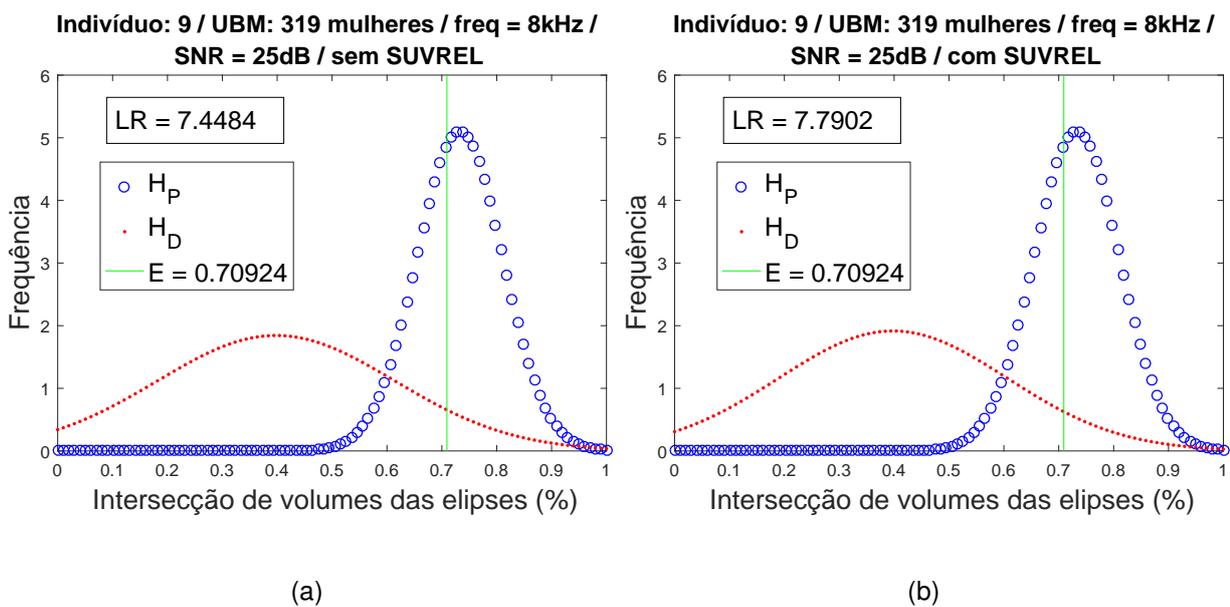
Fonte: produção própria.

Figura A.68. – Cálculo de LR no cenário apresentado no título de cada gráfico.



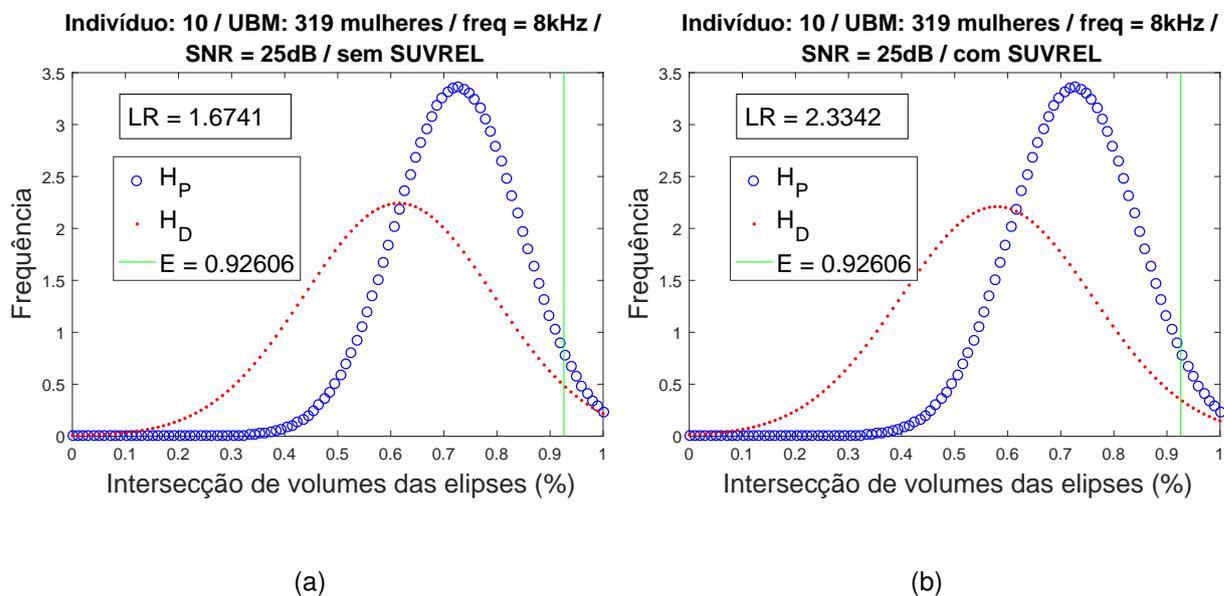
Fonte: produção própria.

Figura A.69. – Cálculo de LR no cenário apresentado no título de cada gráfico.



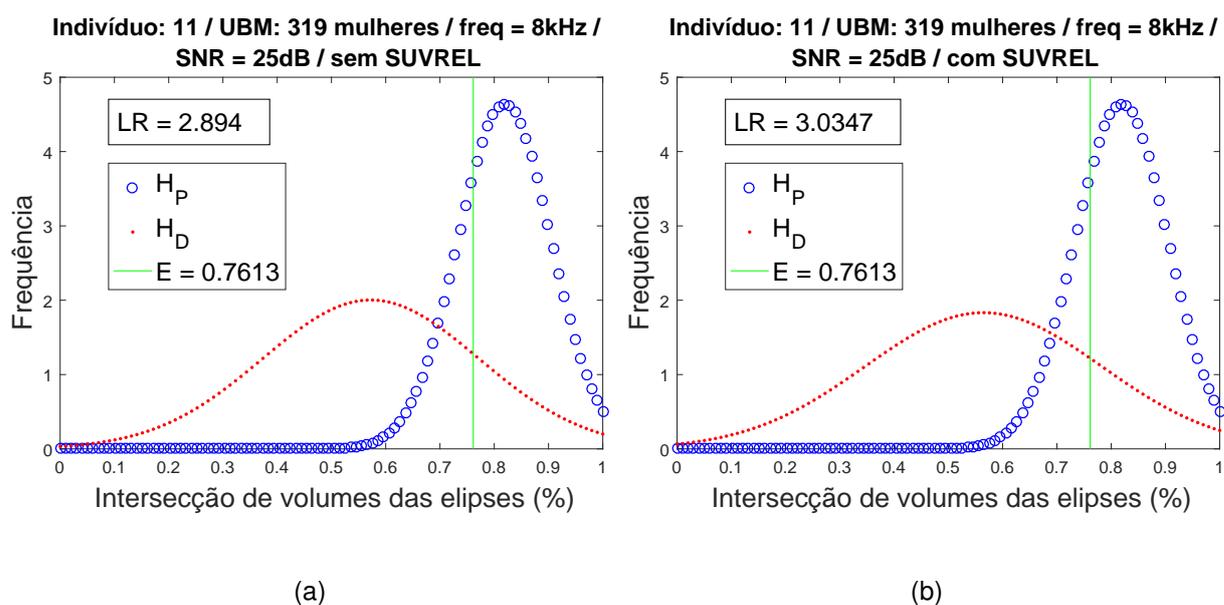
Fonte: produção própria.

Figura A.70. – Cálculo de LR no cenário apresentado no título de cada gráfico.



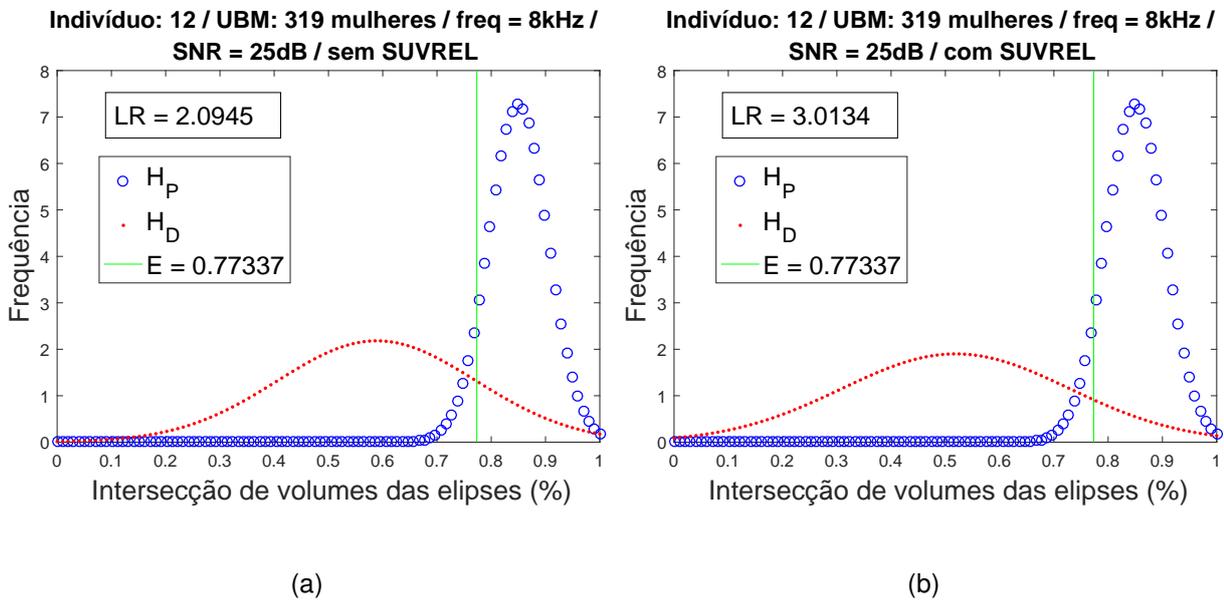
Fonte: produção própria.

Figura A.71. – Cálculo de LR no cenário apresentado no título de cada gráfico.



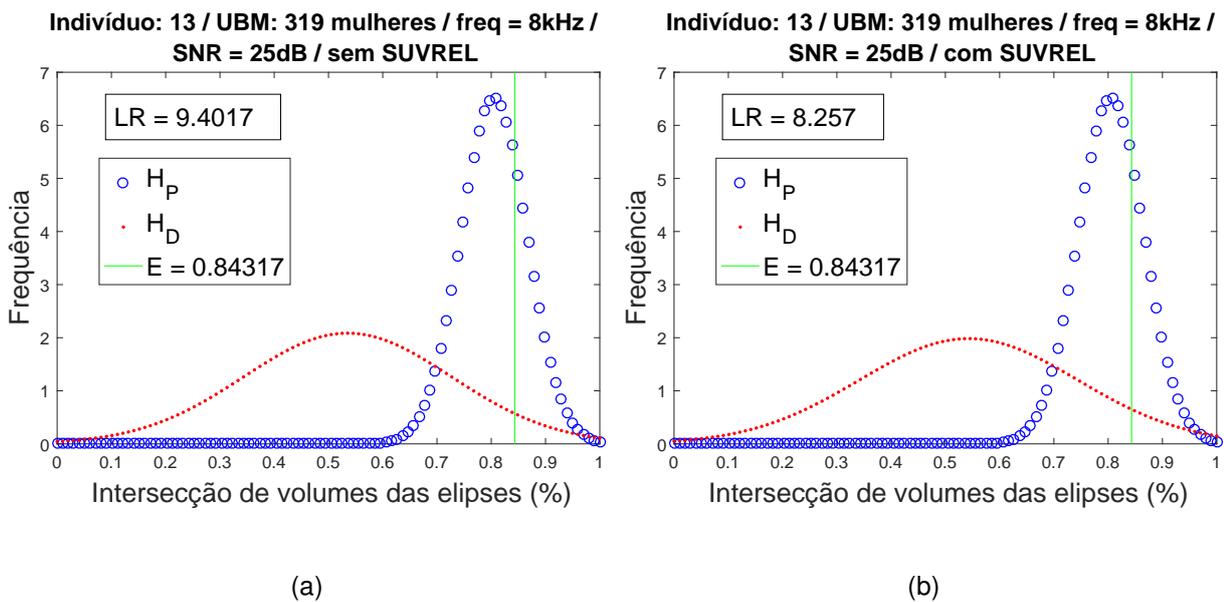
Fonte: produção própria.

Figura A.72. – Cálculo de LR no cenário apresentado no título de cada gráfico.



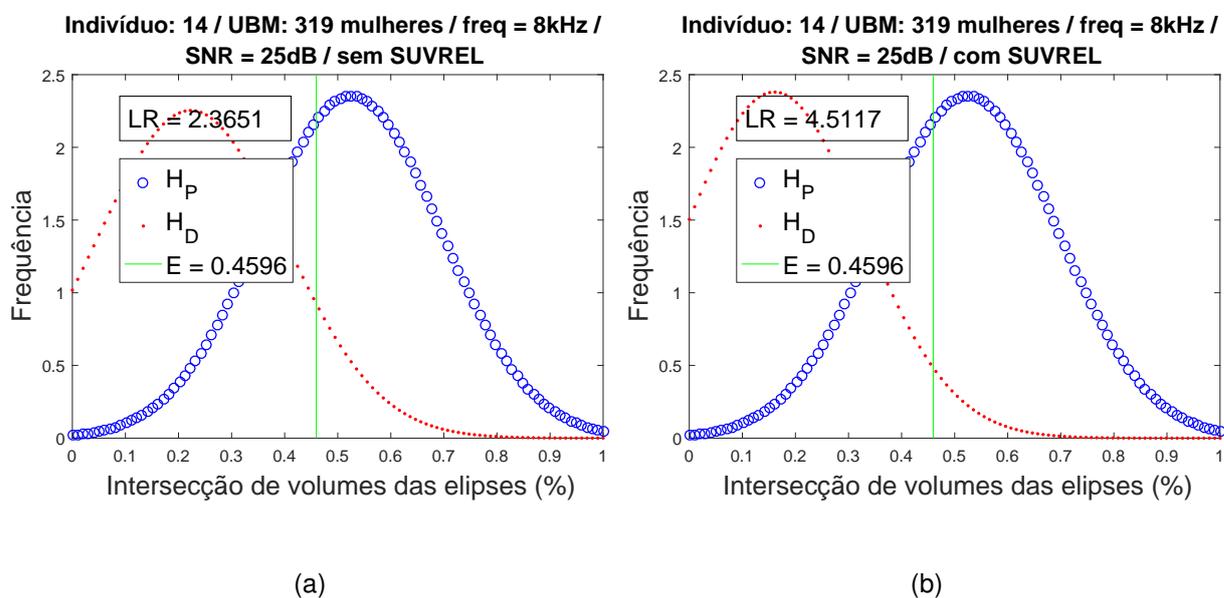
Fonte: produção própria.

Figura A.73. – Cálculo de LR no cenário apresentado no título de cada gráfico.



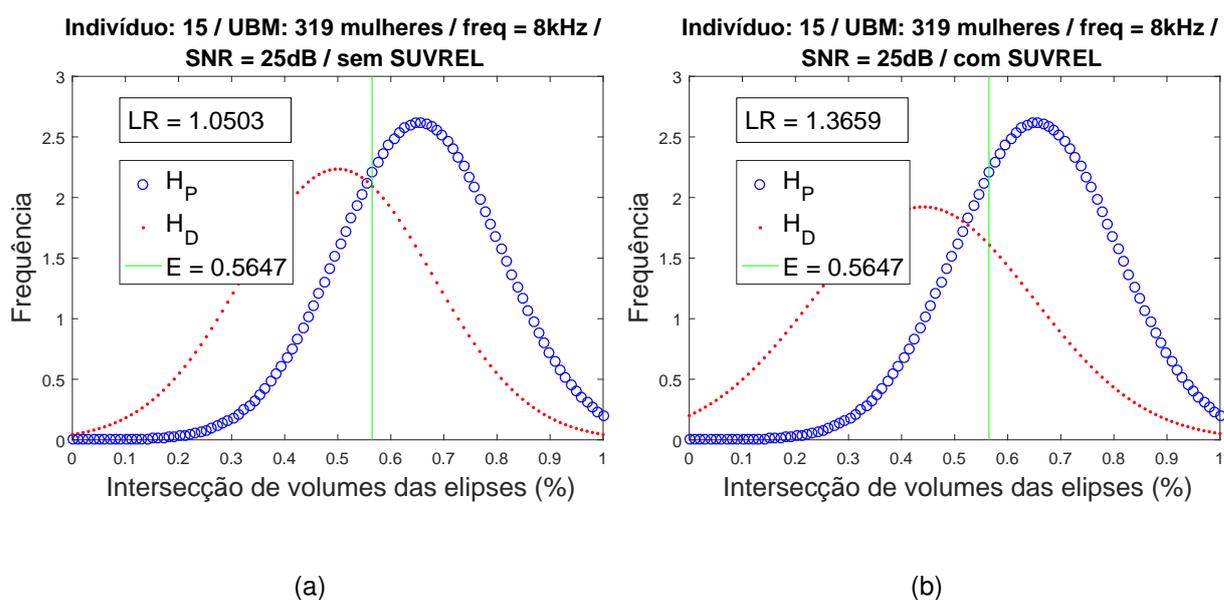
Fonte: produção própria.

Figura A.74. – Cálculo de LR no cenário apresentado no título de cada gráfico.



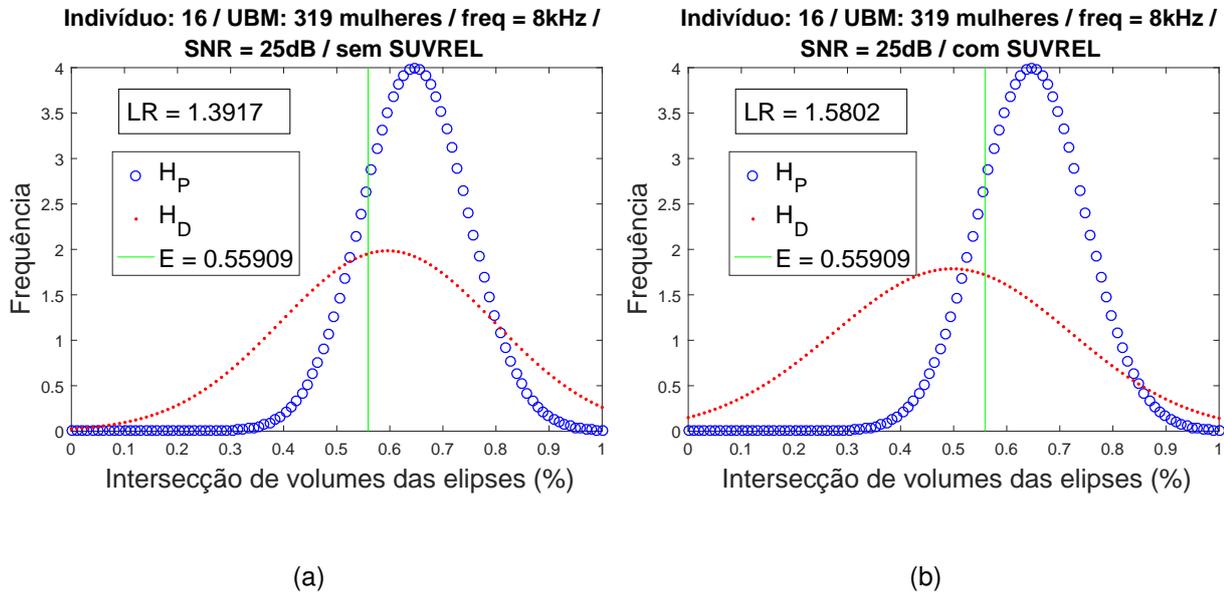
Fonte: produção própria.

Figura A.75. – Cálculo de LR no cenário apresentado no título de cada gráfico.



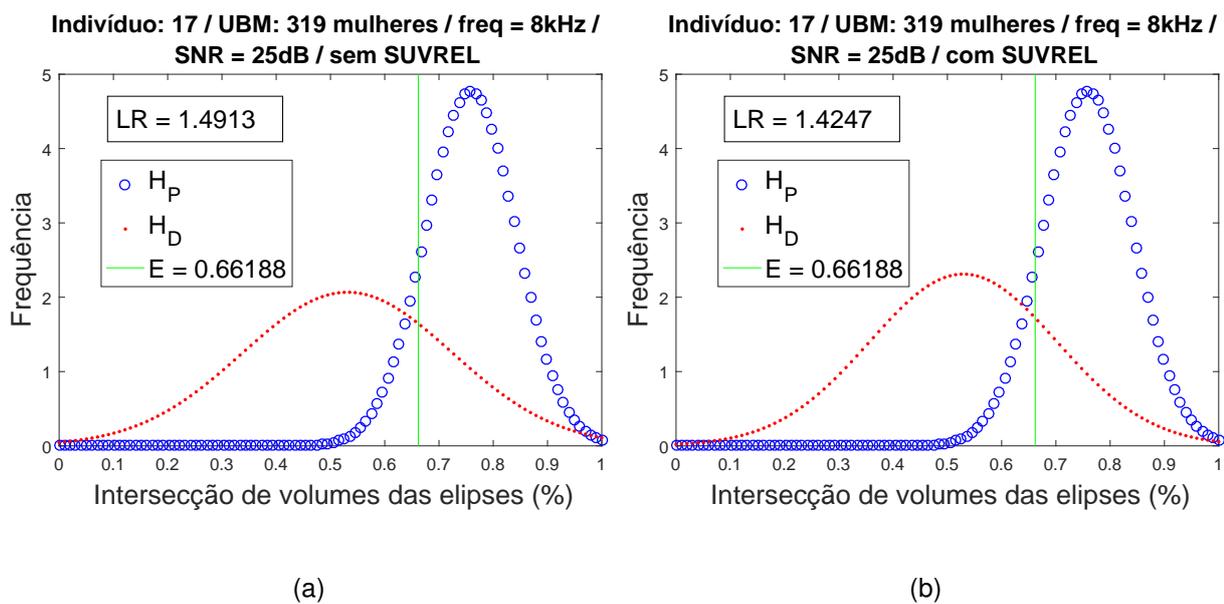
Fonte: produção própria.

Figura A.76. – Cálculo de LR no cenário apresentado no título de cada gráfico.



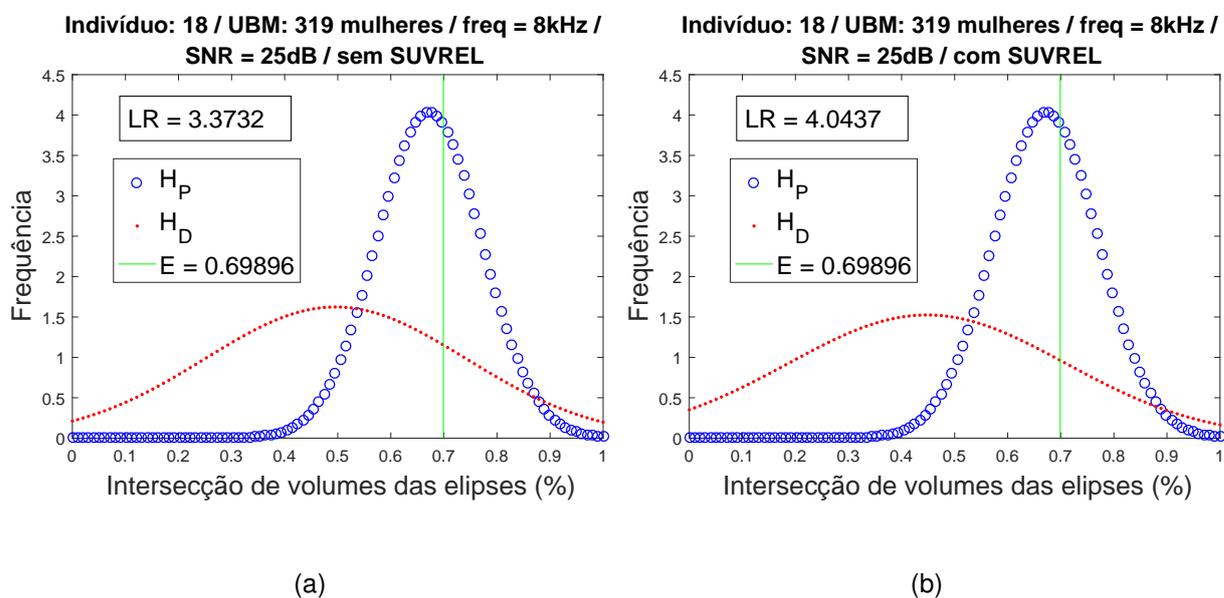
Fonte: produção própria.

Figura A.77. – Cálculo de LR no cenário apresentado no título de cada gráfico.



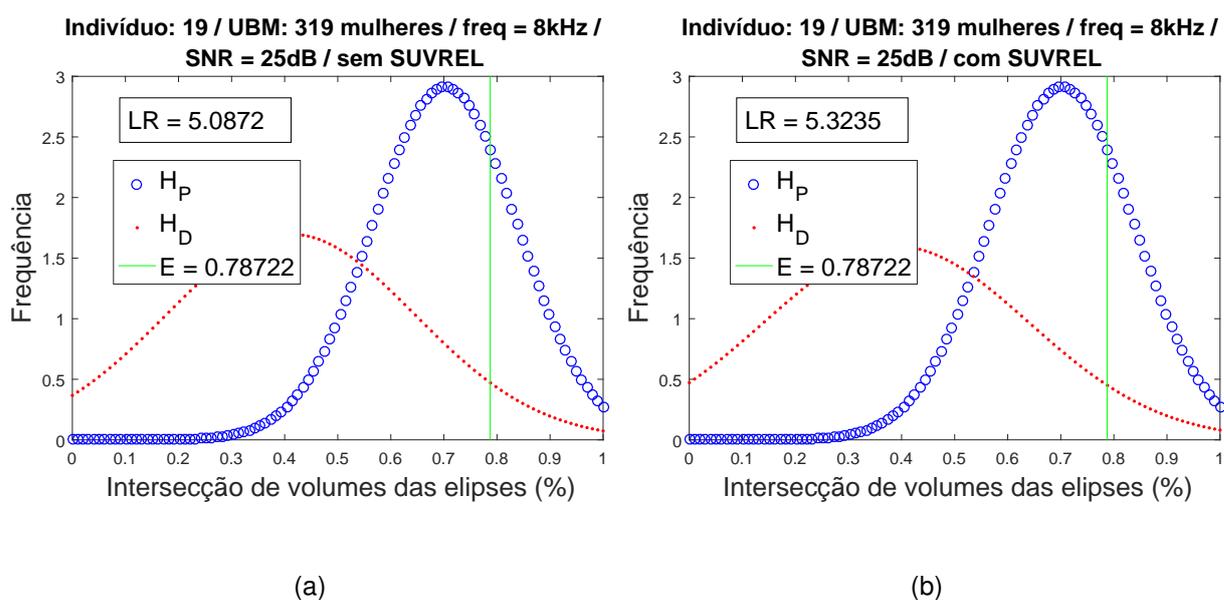
Fonte: produção própria.

Figura A.78. – Cálculo de LR no cenário apresentado no título de cada gráfico.



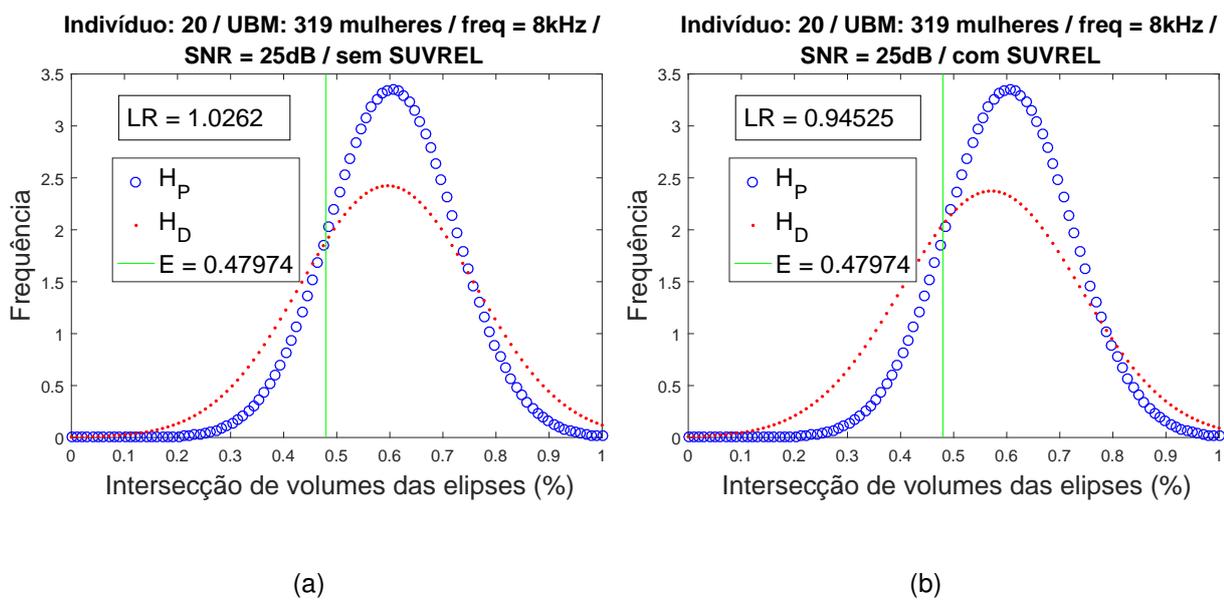
Fonte: produção própria.

Figura A.79. – Cálculo de LR no cenário apresentado no título de cada gráfico.



Fonte: produção própria.

Figura A.80. – Cálculo de LR no cenário apresentado no título de cada gráfico.



Fonte: produção própria.

B. Códigos Matlab

Seguem os códigos MATLAB⁶¹ referentes ao experimento final (seção 5.2.3). Foi usada a versão Matlab R2016a@.

B.1. Script “MFCC_PCA_script_intravariability.m”

```

1 %%
2 %Lists all files provided in a directory organized as:
3 % *each directory represents a subject
4 % *recordings are put in each directory
5
6 close all hidden;
7 fclose all;
8 clear;
9 home;
10
11 %%
12 DIRECTORY_DATA = 'F:\Geoffreys Database\female_cnv_2016
    -08-20\female\cnv\';
13 %DIRECTORY_DATA = 'F:\Geoffreys Database\male_cnv_2016-08-20\
    male\cnv\';
14 DIRECTORY_SAVE = 'F:\SUVREL\FINAL-female-SNR25\';
15 mkdir(DIRECTORY_SAVE);

```

⁶¹Os arquivos estão disponíveis para *download* em <https://drive.google.com/drive/folders/1sq3GxgET6oWvM-vEnmdxmip26QRfao_6?usp=sharing>.

```
16 RECORDING_SESSION = 1;
17
18 list = dir(DIRECTORY_DATA);
19 if ismac == 1
20     list = list(4:end);
21 else
22     list = list(3:end);
23 end
24
25 NUMBER_OF_SUBJECTS = 20; %choose a quantity. If 0, all
    subjects available will be used.
26
27 if NUMBER_OF_SUBJECTS == 0
28     for i=1:size(list,1)
29         if list(i).name(6) == num2str(RECORDING_SESSION)
30             NUMBER_OF_SUBJECTS = NUMBER_OF_SUBJECTS+1;
31         end
32     end
33 end
34
35 SAME_SPEAKER_HYP = [];
36 PLOTS = 0; %change to 1 to get all plots
37
38 for index=1:NUMBER_OF_SUBJECTS
39     DIRECTORY_DATA = 'F:\Geoffreys Database\female_cnv_2016
        -08-20\female\cnv\';
40     cont = 0;
41     EACH_SAMPLE_SIZE = 30; %in seconds
42
43     list_population_temp = [];
44     list_population_temp = dir(DIRECTORY_DATA);
45     list_population = [];
```

```

46
47     if ismac == 1
48         list_population_temp = list_population_temp(4:end);
49     else
50         list_population_temp = list_population_temp(3:end);
51     end
52
53     for i = 1:size(list_population_temp,1)
54         if list_population_temp(i).name(6) == num2str(
55             RECORDING_SESSION)
56             cont = cont + 1;
57             list_population(cont,1).name =
58                 list_population_temp(i).name;
59         end
60         if cont >= NUMBER_OF_SUBJECTS
61             break;
62         end
63     end
64
65     for n = 1:size(list_population,1)
66         list_files{n,1} = dir([DIRECTORY_DATA list_population
67             (n).name]);
68         list_files{n,1} = list_files{n,1}(3:end); %this step
69             eliminates windows directory information
70     end
71
72     USE_DELTAS = 1; %change to 1 to use MFCC deltas and
73         deltadeltas
74
75     SUBJECT = list_population(index).name;
76     DIRECTORY_MFCC = [DIRECTORY_SAVE '
77         dados_cnv_female_MFCC01_30s8000Hz_15DDMFCC_subj_'

```

```

        SUBJECT '_SNR25\'];
72   mkdir([DIRECTORY_MFCC SUBJECT]);
73
74   y = [];
75   Fs = [];
76
77   for file = 1:length(list_files{index})
78       [y_transient, Fs_transient] = audioread([
           DIRECTORY_DATA list_population(index).name '/'
           list_files{index}(file).name]);
79       y = [y ; y_transient];
80       Fs = [Fs ; Fs_transient];
81   end
82
83   if sum(Fs)/Fs(1) ~= length(Fs)
84       display('There is at least one file with a different
           sample rate frequency');
85   else
86       Fs = Fs(1);
87   end
88
89   samples = floor(length(y)/Fs/EACH_SAMPLE_SIZE);
90
91   for i=1:samples
92       mkdir([DIRECTORY_MFCC SUBJECT '\ SUBJECT '_' num2str
           (i)]);
93       audiowrite([DIRECTORY_MFCC SUBJECT '\ SUBJECT '_'
           num2str(i) '\ list_population(index).name '-'
           num2str(EACH_SAMPLE_SIZE) 's-sample' num2str(i) '.
           wav'], y( ((EACH_SAMPLE_SIZE*Fs*(i-1))+1):((
           EACH_SAMPLE_SIZE*Fs*(i))) ), Fs);
94   end

```

```

95
96     DIRECTORY_DATA = [DIRECTORY_MFCC SUBJECT '\'];
97
98     list_population_temp = [];
99     list_population_temp = dir([DIRECTORY_DATA]);
100    list_population = [];
101
102    if ismac == 1
103        list_population_temp = list_population_temp(4:end);
104    else
105        list_population_temp = list_population_temp(3:end);
106    end
107
108    cont = 0;
109
110    USE_DELTAS = 1; %change to 1 to use MFCC deltas and
111                   deltadeltas
112
113    for i = 1:size(list_population_temp,1)
114        if list_population_temp(i).name(6) == num2str(
115            RECORDING_SESSION)
116            cont = cont + 1;
117            list_population(cont,1).name =
118                list_population_temp(i).name;
119        end
120    end
121
122    for n = 1:size(list_population,1)
123        list_files{n,1} = dir([DIRECTORY_DATA list_population
124            (n).name]);
125        list_files{n,1} = list_files{n,1}(3:end); %this step
126            eliminates windows directory information

```

```
122     end
123
124     %How to refer to these cell contents:
125     % * list_population is an struct array with the following
        fields for each directory:
126     %     name
127     %     date
128     %     bytes
129     %     isdir
130     %     datenum
131     % E.g.: to refer to the 10th directory name use:
132     %     list_population(10).name
133     %
134     % * list_files is a cell with one struct array with
        fields (as above) listing all files saved in each
        directory
135     % E.g.: to refer to the 23rd file in the 10th directory
        use:
136     %     list_files{10}(23).name
137
138     RESAMPLEAUDIO = 8000; %0 = don't resample. To resample,
        change it to the desired sample value in Hz.
139
140     ADDNOISE = 1; %0 = don't introduce noise.
141     SNR = 25; %Signal to Noise Ratio with SNR=20 log10(
        signal_power / noise_power )
142
143     for subj = 1:length(list_population)
144         y = [];
145         Fs = [];
146
147         for file = 1:length(list_files{subj})
```

```

148     [y_transient, Fs_transient] = audioread([
        DIRECTORY_DATA list_population(subj).name '/'
        list_files{subj}(file).name]);
149     y = [y ; y_transient];
150     Fs = [Fs ; Fs_transient];
151     if length(y) >= EACH_SAMPLE_SIZE * Fs(1)
152         break;
153     end
154 end

155
156 if sum(Fs)/Fs(1) ~= length(Fs)
157     display('There is at least one file with a
        different sample rate frequency');
158 else
159     Fs = Fs(1);
160 end
161 audiowrite([DIRECTORY_MFCC list_population(subj).name
        '-' num2str(EACH_SAMPLE_SIZE) 's-sample.wav'], y,
        Fs);
162
163 if RESAMPLEAUDIO ~= 0
164     [P,Q] = rat(RESAMPLEAUDIO/Fs);
165     y_new = resample(y,P,Q);
166     y = y_new;
167     Fs = RESAMPLEAUDIO;
168     audiowrite([DIRECTORY_MFCC list_population(subj).
        name '-' num2str(EACH_SAMPLE_SIZE) 's-sample'
        num2str(RESAMPLEAUDIO) '.wav'], y, Fs);
169 end
170
171 if ADDNOISE ~= 0
172     Z = awgn(y, SNR, 'measured');

```

```

173     y = Z;
174     if RESAMPLEAUDIO ~= 0
175         audiowrite([DIRECTORY_MFCC list_population(
                subj).name '-' num2str(EACH_SAMPLE_SIZE) '
                s-sample' num2str(RESAMPLEAUDIO) '-SNR'
                num2str(SNR) '.wav'], y, Fs);
176     else
177         audiowrite([DIRECTORY_MFCC list_population(
                subj).name '-' num2str(EACH_SAMPLE_SIZE) '
                s-sample-SNR' num2str(SNR) '.wav'], y, Fs)
                ;
178     end
179 end
180
181 FRAME = ':';
182 framelength = 0.025;
183 framestep = 0.010;
184 samplelength = floor(framelength * Fs);
185 samplestep = floor(framestep * Fs);
186 ntotalsamples = size(y,1);
187
188 %This stage guarantees that the total number of
        samples is a multiple of
189 %the samplestep by adding the necessary amount of
        zeros at the end
190 if mod(ntotalsamples, samplestep) ~= 0
191     y = [y; zeros(ceil(ntotalsamples/samplestep)*
                samplestep - ntotalsamples, 1)];
192     ntotalsamples = size(y,1);
193 end
194
195 sampleposition = 1;

```

```

196     i = 1;
197     signal_frames = [];
198
199     while (sampleposition + samplelength) <=
200         ntotalsamples
201         signal_frames(:,i) = y(sampleposition:(
202             sampleposition+samplelength-1),1);
203         sampleposition = sampleposition + samplestep;
204         i = i+1;
205     end
206
207     signal = signal_frames(:,FRAME);
208
209     window = hamming(size(signal,1));
210     pxx = [];
211     for l = 1: size(signal,2)
212         [pxx_temp,f] = periodogram(signal(:,l),window,
213             size(signal,1),Fs);
214         pxx = [pxx pxx_temp];
215     end
216
217     %%%%%%%%%%%
218
219     n_triangular_filters = 15;
220     N_POINTS = n_triangular_filters-2;
221
222     if subj==1
223         [H, hi, mi] = melTriangularFilterBank(
224             n_triangular_filters, samplelength, 0, Fs);
225         %xlim([0 , 4000]);
226     else

```

```

223         [H, hi, mi] = melTriangularFilterBank(
                n_triangular_filters, samplelength, 0, Fs);
224     end
225
226     %MFCC:
227     melscaledlogenergy = zeros(size(pxx,2),
                n_triangular_filters);
228
229     for sample=1:size(pxx,2)
230         for j = 1:n_triangular_filters
231             %filter application
232             melscaledlogenergy(sample, j) = sum(pxx(:,
                sample).*H(:,j));
233         end
234     end
235
236     %%%%%%%%%%%
237
238     x = hi(2:end-1);
239     newx = hi(1,1):hi(1, end);
240     y = log10(melscaledlogenergy);
241     newy = spline(x,y,newx); %Complete spline
                interpolation
242
243     MFCCtemp = zeros(size(pxx,2), n_triangular_filters);
244
245     for sample = 1:size(pxx,2)
246         MFCCtemp(sample,:) = dct(y(sample,:));
247     end
248
249     deltaMFCC = zeros(size(pxx,2), n_triangular_filters
                -4);

```

```

250     deltadeltaMFCC = zeros(size(pxx,2),
251         n_triangular_filters-6);
252
253     %deltas and delta-deltas
254     for sample = 1:size(pxx,2)
255         for k = 3:size(MFCCtemp,2)-2
256             deltaMFCC(sample,k-2) = ((MFCCtemp(sample, k
257                 +2) - MFCCtemp(sample, k-2)) + (MFCCtemp(
258                 sample, k+1) - MFCCtemp(sample, k-1)))/5;
259         end
260     end
261
262     for sample = 1:size(pxx,2)
263         for k = 3:size(deltaMFCC,2)-2
264             deltadeltaMFCC(sample,k-2) = ((deltaMFCC(
265                 sample, k+2) - deltaMFCC(sample, k-2)) + (
266                 deltaMFCC(sample, k+1) - deltaMFCC(sample,
267                 k-1)))/5;
268         end
269     end
270
271     if USE_DELTAS == 1
272         MFCC = [MFCCtemp(:,2:N_POINTS+1) deltaMFCC(:,2:
273             end) deltadeltaMFCC(:,2:end)];%with deltas
274     else
275         MFCC = [MFCCtemp(:,2:N_POINTS+1)];%without deltas
276     end
277
278     %figure;
279     % stem(MFCC(10,:));
280     % ax = gca;
281     % ax.XAxis.Exponent = 0;

```

```

275     % ax.XAxis.TickLabelFormat = '%6.0f';
276     % title('MFCC');
277     % xlabel('Coefficient Index');
278     % ylabel('Amplitude');
279     % saveas(gcf, [DIREC NOME '-MFCC'], 'epsc');
280     % saveas(gcf, [DIREC NOME '-MFCC'], 'jpg');
281
282     save([DIRECTORY_MFCC list_population(subj).name '-
           dados-MFCC'] , 'MFCC');
283
284 end
285
286 % WITHOUT SUVREL
287
288 PLOTFACOR = 4; %Reduction factor acting on radius'
           standard deviation just to plot.
289
290 dadosParaPCA = [];
291 categorias = {'inicio' 0};
292
293 for subj = 1:length(list_population)
294     load([DIRECTORY_MFCC list_population(subj).name '-
           dados-MFCC']);
295     i = isnan(MFCC);
296     j = find(i(:,2)~=1);
297     dadosParaPCA = [dadosParaPCA; MFCC(j,1:end)];
298     categorias{subj+1,1} = list_population(subj).name;
299     categorias{subj+1,2} = categorias{subj,2}+ length(j);
300 end
301
302 [coeff, score, latent, tsquared, explained, mu] = pca(
           dadosParaPCA);

```

```
303
304     if PLOTS == 1
305         figure;
306         %limits:
307         MIN1 = min([score(:,1); score(:,2); score(:,3)]);
308         MIN2 = MIN1;
309         MIN3 = MIN1;
310         MAX1 = max([score(:,1); score(:,2); score(:,3)]);
311         MAX2 = MAX1;
312         MAX3 = MAX1;
313
314         subplot(3,3,1), plot(score(:,1), score(:,1), '.r')
315         xlim([MIN1 MAX1]); ylim([MIN1 MAX1]);
316         ylabel('PCA1');
317
318         subplot(3,3,2), plot(score(:,2), score(:,1), '.r')
319         xlim([MIN2 MAX2]); ylim([MIN1 MAX1]);
320
321         subplot(3,3,3), plot(score(:,3), score(:,1), '.r')
322         xlim([MIN3 MAX3]); ylim([MIN1 MAX1]);
323
324         subplot(3,3,4), plot(score(:,1), score(:,2), '.r')
325         xlim([MIN1 MAX1]); ylim([MIN2 MAX2]);
326         ylabel('PCA2');
327
328         subplot(3,3,5), plot(score(:,2), score(:,2), '.r')
329         xlim([MIN2 MAX2]); ylim([MIN2 MAX2]);
330
331         subplot(3,3,6), plot(score(:,3), score(:,2), '.r')
332         xlim([MIN3 MAX3]); ylim([MIN2 MAX2]);
333
334         subplot(3,3,7), plot(score(:,1), score(:,3), '.r')
```

```
335     xlim([MIN1 MAX1]); ylim([MIN3 MAX3]);
336     xlabel('PCA1'); ylabel('PCA3');
337
338     subplot(3,3,8), plot(score(:,2), score(:,3), '.r')
339     xlim([MIN2 MAX2]); ylim([MIN3 MAX3]);
340     xlabel('PCA2');
341
342     subplot(3,3,9), plot(score(:,3), score(:,3), '.r')
343     xlim([MIN3 MAX3]); ylim([MIN3 MAX3]);
344     xlabel('PCA3');
345
346     figure;
347     for h = 1:length(list_population)
348         cor = rem(h,15);
349
350         if cor==1
351             pontos = [1 0 0];
352         elseif cor ==2
353             pontos = [0 0 1];
354         elseif cor ==3
355             pontos = [0 1 0];
356         elseif cor ==4
357             pontos = [0 0 0];
358         elseif cor ==5
359             pontos = [1 0 1];
360         elseif cor ==6
361             pontos = [1 1 0];
362         elseif cor ==7
363             pontos = [0 1 1];
364         elseif cor ==8
365             pontos = [0.5 0.5 1];
366         elseif cor ==9
```

```
367         pontos = [0.5 1 0.5];
368     elseif cor ==10
369         pontos = [0 0.5 0];
370     elseif cor ==11
371         pontos = [0.5 0 0.5];
372     elseif cor ==12
373         pontos = [0.5 0.5 0];
374     elseif cor ==13
375         pontos = [0.8 1 0.2];
376     elseif cor ==14
377         pontos = [0.2 0.6 0.7];
378     elseif cor ==0
379         pontos = [0.3 0.9 0.1];
380     end
381
382     plot3(score(categorias{h,2}+1:categorias{h
383         +1,2},1), ...
384         score(categorias{h,2}+1:categorias{h+1,2},2),
385         ...
386         score(categorias{h,2}+1:categorias{h+1,2},3)
387         ,...
388         'Marker', '.', 'LineStyle', 'none', 'Color',
389         pontos);
390     hold on;
391 end
392 grid on;
393 %title('SEM SUVREL');
394 xlabel('PCA 1');
395 ylabel('PCA 2');
396 zlabel('PCA 3');
397 % legend('subject A', 'subject B', 'subject C', '
398     subject D', 'subject E', ...
```

```

394         %      'subject F', 'subject G', 'subject H', 'subject
           I', 'subject J', ...
395         %      'subject K', 'subject L', 'subject M', 'subject
           N', 'subject O');
396         hold off;
397     end
398
399     for h = 1:length(list_population)
400         raiosX(h,:) = std(score(categorias{h,2}+1:categorias{
           h+1,2},1:3));
401         centrosX(h,:) = mean(score(categorias{h,2}+1:
           categorias{h+1,2},1:3));
402     end
403
404     if PLOTS == 1
405         figure;
406
407         [x, y, z] = sphere;
408         for h = 1:length(list_population)
409             cor = rem(h,15);
410
411             if cor==1
412                 superf = [1 0 0];
413             elseif cor ==2
414                 superf = [0 0 1];
415             elseif cor ==3
416                 superf = [0 1 0];
417             elseif cor ==4
418                 superf = [0 0 0];
419             elseif cor ==5
420                 superf = [1 0 1];
421             elseif cor ==6

```

```
422         superf = [1 1 0];
423     elseif cor ==7
424         superf = [0 1 1];
425     elseif cor ==8
426         superf = [0.5 0.5 1];
427     elseif cor ==9
428         superf = [0.5 1 0.5];
429     elseif cor ==10
430         superf = [0 0.5 0];
431     elseif cor ==11
432         superf = [0.5 0 0.5];
433     elseif cor ==12
434         superf = [0.5 0.5 0];
435     elseif cor ==13
436         superf = [0.8 1 0.2];
437     elseif cor ==14
438         superf = [0.2 0.6 0.7];
439     elseif cor ==0
440         superf = [0.3 0.9 0.1];
441     end
442
443     surf(x*(raiosX(h, 1)./PLOTFACTOR)+centrosX(h, 1),
         y*(raiosX(h, 2)./PLOTFACTOR)+centrosX(h, 2),
         z*(raiosX(h, 3)./PLOTFACTOR)+centrosX(h, 3), '
         FaceAlpha',0.2,'LineStyle','none', 'FaceColor'
         , superf );
444     hold on;
445     end
446
447     grid on;
448     title('SEM SUVREL');
449     xlabel('PCA 1');
```

```
450     ylabel('PCA 2');
451     zlabel('PCA 3');
452
453     % legend('subject A', 'subject B', 'subject C', '
         subject D', 'subject E', ...
454     %       'subject F', 'subject G', 'subject H', 'subject
         I', 'subject J', ...
455     %       'subject K', 'subject L', 'subject M', 'subject
         N', 'subject O');
456
457     hold off;
458
459     figure;
460     hold on;
461     t=-pi:0.01:pi;
462     for h = 1:length(list_population)
463         x0 = centrosX(h,1);
464         y0 = centrosX(h,2);
465         rx = (raiosX(h,1)./PLOTFACTOR);
466         ry = (raiosX(h,2)./PLOTFACTOR);
467         x = x0 + rx*cos(t);
468         y = y0 + ry*sin(t);
469         cor = rem(h,15);
470
471         if cor==1
472             pontos = [1 0 0];
473         elseif cor ==2
474             pontos = [0 0 1];
475         elseif cor ==3
476             pontos = [0 1 0];
477         elseif cor ==4
478             pontos = [0 0 0];
```

```
479         elseif cor ==5
480             pontos = [1 0 1];
481         elseif cor ==6
482             pontos = [1 1 0];
483         elseif cor ==7
484             pontos = [0 1 1];
485         elseif cor ==8
486             pontos = [0.5 0.5 1];
487         elseif cor ==9
488             pontos = [0.5 1 0.5];
489         elseif cor ==10
490             pontos = [0 0.5 0];
491         elseif cor ==11
492             pontos = [0.5 0 0.5];
493         elseif cor ==12
494             pontos = [0.5 0.5 0];
495         elseif cor ==13
496             pontos = [0.8 1 0.2];
497         elseif cor ==14
498             pontos = [0.2 0.6 0.7];
499         elseif cor ==0
500             pontos = [0.3 0.9 0.1];
501         end
502         plot(x,y, 'Color', pontos);
503     end
504     hold off;
505 end
506
507 ab_vol = zeros(length(list_population), 1);
508 shared_volume = zeros(length(list_population));
509
510 for i=1:length(list_population)
```

```

511     for j=i:length(list_population)
512         [abv1, abv2, sv12, sv21] =
                    ellipsoidVolumeComparison(...
513             centrosX(i,:), centrosX(j,:), raiosX(i,:),
                    raiosX(j,:), 5e5);
514         ab_vol(i, 1) = abv1;
515         if i==j
516             shared_volume(i, j) = sv12;
517         else
518             shared_volume(i, j) = sv12;
519             shared_volume(j, i) = sv21;
520         end
521     end
522 end
523
524 if PLOTS == 1
525     figure1 = figure;
526     colormap('jet');
527     axes1 = axes('Parent',figure1);
528     hold(axes1,'on');
529     image(shared_volume, 'Parent', axes1, 'CDataMapping', '
                    scaled');
530     xlim(axes1, [0.5 length(list_population)+0.5]);
531     ylim(axes1, [0.5 length(list_population)+0.5]);
532     box(axes1, 'on');
533     axis(axes1, 'ij');
534     set(axes1, 'Layer', 'top');
535     colorbar('peer', axes1);
536 end
537
538 SAME_SPEAKER_HYP(index, 1) = mean(shared_volume(1, 2:end));
539 SAME_SPEAKER_HYP(index, 2) = std(shared_volume(1, 2:end));

```

```

540
541     EVIDENCE(index,1) = shared_volume(1,3);
542 end
543
544 save([DIRECTORY_SAVE 'SAME_SPEAKER_HYP'], 'SAME_SPEAKER_HYP')
    ;
545 save([DIRECTORY_SAVE 'EVIDENCE'], 'EVIDENCE');

```

B.2. Script “MFCC_PCA_script_UBM_SUVREL.m”

```

1 %%
2 %Lists all files provided in a directory organized as:
3 % *each directory represents a subject
4 % *recordings are put in each directory
5
6 close all hidden;
7 fclose all;
8 clear;
9 home;
10 %
11 %DIRECTORY_DATA = 'F:\Geoffreys Database\female_cnv_2016
    -08-20\female\cnv\';
12 DIRECTORY_DATA = 'F:\Geoffreys Database\male_cnv_2016-08-20\
    male\cnv\';
13
14 DIRECTORY_SAVE = 'F:\SUVREL\FINAL-female-SNR25\';
15
16 GENDER = 'mulheres';
17
18 DIRECTORY_MFCC = [DIRECTORY_SAVE '
    dados_cnv_female_MFCC01_30s8000Hz_15MFCC_319subj_SNR25\'];
19 mkdir(DIRECTORY_MFCC);

```

```
20
21 EACH_SAMPLE_SIZE = 30; %in seconds
22
23 RESAMPLEAUDIO = 8000; %0 = don't resample. To resample,
    change it to the desired sample value in Hz.
24
25 ADDNOISE = 1; %0 = don't introduce noise.
26 SNR = 25; %Signal to Noise Ratio with SNR=20 log10(
    signal_power / noise_power )
27
28 list_population_temp = dir(DIRECTORY_DATA);
29
30 if ismac == 1
31     list_population_temp = list_population_temp(4:end);
32 else
33     list_population_temp = list_population_temp(3:end);
34 end
35
36 RECORDING_SESSION = 1;
37 NUMBER_OF_SUBJECTS = 319;
38 cont = 0;
39
40 USE_DELTAS = 1; %change to 1 to use MFCC deltas and
    deltadeltas
41
42 for i = 1:size(list_population_temp,1)
43     if list_population_temp(i).name(6) == num2str(
        RECORDING_SESSION)
44         cont = cont + 1;
45         list_population(cont,1).name = list_population_temp(i
            ).name;
46     end
```

```
47
48     if cont >= NUMBER_OF_SUBJECTS
49         break;
50     end
51 end
52
53 for n = 1:size(list_population,1)
54     list_files{n,1} = dir([DIRECTORY_DATA list_population(n).
55         name]);
56     list_files{n,1} = list_files{n,1}(3:end); %this step
57         eliminates windows directory information
58 end
59
60 %How to refer to these cell contents:
61 % * list_population is an struct array with the following
62     fields for each directory:
63 %     name
64 %     date
65 %     bytes
66 %     isdir
67 %     datenum
68 % E.g.: to refer to the 10th directory name use:
69 %     list_population(10).name
70 %
71 % * list_files is a cell with one struct array with fields (
72     as above) listing all files saved in each directory
73 % E.g.: to refer to the 23rd file in the 10th directory use:
74 %     list_files{10}(23).name
75
76 %%
77
78 for subj = 1:length(list_population)
```

```

75     y = [];
76     Fs = [];
77
78     for file = 1:length(list_files{subj})
79         [y_transient, Fs_transient] = audioread([
            DIRECTORY_DATA list_population(subj).name '/'
            list_files{subj}(file).name]);
80         y = [y ; y_transient];
81         Fs = [Fs ; Fs_transient];
82         if length(y) >= EACH_SAMPLE_SIZE * Fs(1)
83             break;
84         end
85     end
86
87     if sum(Fs)/Fs(1) ~= length(Fs)
88         display('There is at least one file with a different
            sample rate frequency');
89     else
90         Fs = Fs(1);
91     end
92     audiowrite([DIRECTORY_MFCC list_population(subj).name '-
            num2str(EACH_SAMPLE_SIZE) 's-sample.wav'], y, Fs);
93
94     if RESAMPLEAUDIO ~= 0
95         [P,Q] = rat(RESAMPLEAUDIO/Fs);
96         y_new = resample(y,P,Q);
97         y = y_new;
98         Fs = RESAMPLEAUDIO;
99         audiowrite([DIRECTORY_MFCC list_population(subj).name
            '- num2str(EACH_SAMPLE_SIZE) 's-sample' num2str(
            RESAMPLEAUDIO) '.wav'], y, Fs);
100    end

```

```
101
102     if SNR > 0
103         if ADDNOISE ~= 0
104             Z = awgn(y, SNR, 'measured');
105             y = Z;
106             if RESAMPLEAUDIO ~= 0
107                 audiowrite([DIRECTORY_MFCC list_population(
108                     subj).name '-' num2str(EACH_SAMPLE_SIZE) '
109                     s-sample' num2str(RESAMPLEAUDIO) '-SNR'
110                     num2str(SNR) '.wav'], y, Fs);
111             else
112                 audiowrite([DIRECTORY_MFCC list_population(
113                     subj).name '-' num2str(EACH_SAMPLE_SIZE) '
114                     s-sample-SNR' num2str(SNR) '.wav'], y, Fs)
115                 ;
116             end
117         end
118     end
119
120     FRAME = ':';
121
122     framelength = 0.025;
123     framestep = 0.010;
124
125     samplelength = floor(framelength * Fs);
126     samplestep = floor(framestep * Fs);
127
128     ntotalsamples = size(y,1);
129
130     %This stage guarantees that the total number of samples
131     %is a multiple of
132     %the samplestep by adding the necessary amount of zeros
```

```

    at the end
126 if mod(ntotalsamples, samplestep) ~= 0
127     y = [y; zeros(ceil(ntotalsamples/samplestep)*
        samplestep - ntotalsamples, 1)];
128     ntotalsamples = size(y,1);
129 end
130
131 sampleposition = 1;
132 i = 1;
133 signal_frames = [];
134
135 while (sampleposition + samplelength) <= ntotalsamples
136     signal_frames(:,i) = y(sampleposition:(sampleposition
        +samplelength-1),1);
137     sampleposition = sampleposition + samplestep;
138     i = i+1;
139 end
140
141 signal = signal_frames(:,FRAME);
142
143 window = hamming(size(signal,1));
144 pxx = [];
145 for l = 1: size(signal,2)
146     [pxx_temp,f] = periodogram(signal(:,l),window,size(
        signal,1),Fs);
147     pxx = [pxx pxx_temp];
148 end
149 %%%%%%%%%%%
150
151 n_triangular_filters = 15;
152 N_POINTS = n_triangular_filters-2;

```

```

154     if subj==1
155         [H, hi, mi] = melTriangularFilterBank(
                n_triangular_filters, samplelength, 0, Fs);
156         %xlim([0 , 4000]);
157     else
158         [H, hi, mi] = melTriangularFilterBank(
                n_triangular_filters, samplelength, 0, Fs);
159     end
160
161     %MFCC:
162     melscaledlogenergy = zeros(size(pxx,2),
                n_triangular_filters);
163
164     for sample=1:size(pxx,2)
165         for j = 1:n_triangular_filters
166             %filter application
167             melscaledlogenergy(sample, j) = sum(pxx(:,sample)
                .*H(:,j));
168         end
169     end
170
171     %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
172     x = hi(2:end-1);
173     newx = hi(1,1):hi(1, end);
174     y = log10(melscaledlogenergy);
175     newy = spline(x,y,newx); %Complete spline interpolation
176     MFCCtemp = zeros(size(pxx,2), n_triangular_filters);
177
178     for sample = 1:size(pxx,2)
179         MFCCtemp(sample,:) = dct(y(sample,:));
180     end
181

```

```

182     deltaMFCC = zeros(size(pxx,2), n_triangular_filters-4);
183     deltadeltaMFCC = zeros(size(pxx,2), n_triangular_filters
      -6);
184
185     %deltas and delta-deltas
186     for sample = 1:size(pxx,2)
187         for k = 3:size(MFCCtemp,2)-2
188             deltaMFCC(sample,k-2) = ((MFCCtemp(sample, k+2) -
      MFCCtemp(sample, k-2)) + (MFCCtemp(sample, k
      +1) - MFCCtemp(sample, k-1)))/5;
189         end
190     end
191
192     for sample = 1:size(pxx,2)
193         for k = 3:size(deltaMFCC,2)-2
194             deltadeltaMFCC(sample,k-2) = ((deltaMFCC(sample,
      k+2) - deltaMFCC(sample, k-2)) + (deltaMFCC(
      sample, k+1) - deltaMFCC(sample, k-1)))/5;
195         end
196     end
197
198     if USE_DELTAS == 1
199         MFCC = [MFCCtemp(:,2:N_POINTS+1) deltaMFCC(:,2:end)
      deltadeltaMFCC(:,2:end)];%with deltas
200     else
201         MFCC = [MFCCtemp(:,2:N_POINTS+1)];%without deltas
202     end
203
204     %figure;
205     % stem(MFCC(10,:));
206     % ax = gca;
207     % ax.XAxis.Exponent = 0;

```

```

208     % ax.XAxis.TickLabelFormat = '%6.0f';
209     % title('MFCC');
210     % xlabel('Coefficient Index');
211     % ylabel('Amplitude');
212     % saveas(gcf, [DIREC NOME '-MFCC'], 'epsc');
213     % saveas(gcf, [DIREC NOME '-MFCC'], 'jpg');
214
215     save([DIRECTORY_MFCC list_population(subj).name '-dados -
          MFCC'] , 'MFCC');
216 end
217
218 %%
219
220 % WITHOUT SUVREL
221
222 PLOTFACOR = 4; %Reduction factor acting on radius' standard
          deviation just to plotting.
223
224 dadosParaPCA = [];
225 categorias = {'inicio' 0};
226
227 for subj = 1:length(list_population)
228     load([DIRECTORY_MFCC list_population(subj).name '-dados -
          MFCC']);
229
230     i = isnan(MFCC);
231     j = find(i(:,2)~=1);
232     dadosParaPCA = [dadosParaPCA; MFCC(j,1:end)];
233
234     categorias{subj+1,1} = list_population(subj).name;
235     categorias{subj+1,2} = categorias{subj,2}+ length(j);
236 end

```

```
237
238 [coeff, score, latent, tsquared, explained, mu] = pca(
    dadosParaPCA);
239
240 % figure;
241 % plot(1:1:length(latent),latent, '.b');
242 %fig2plotly(gcf,'strip',false);
243
244 figure;
245 %limits:
246 MIN1 = min([score(:,1); score(:,2); score(:,3)]);
247 MIN2 = MIN1;
248 MIN3 = MIN1;
249 MAX1 = max([score(:,1); score(:,2); score(:,3)]);
250 MAX2 = MAX1;
251 MAX3 = MAX1;
252
253 % MIN1 = min(score(:,1));
254 % MIN2 = min(score(:,2));
255 % MIN3 = min(score(:,3));
256 % MAX1 = max(score(:,1));
257 % MAX2 = max(score(:,2));
258 % MAX3 = max(score(:,3));
259
260 subplot(3,3,1), plot(score(:,1), score(:,1), '.r')
261 xlim([MIN1 MAX1]); ylim([MIN1 MAX1]);
262 ylabel('PCA1');
263
264 subplot(3,3,2), plot(score(:,2), score(:,1), '.r')
265 xlim([MIN2 MAX2]); ylim([MIN1 MAX1]);
266
267 subplot(3,3,3), plot(score(:,3), score(:,1), '.r')
```

```
268 xlim([MIN3 MAX3]); ylim([MIN1 MAX1]);
269
270 subplot(3,3,4), plot(score(:,1), score(:,2), '.r')
271 xlim([MIN1 MAX1]); ylim([MIN2 MAX2]);
272 ylabel('PCA2');
273
274 subplot(3,3,5), plot(score(:,2), score(:,2), '.r')
275 xlim([MIN2 MAX2]); ylim([MIN2 MAX2]);
276
277 subplot(3,3,6), plot(score(:,3), score(:,2), '.r')
278 xlim([MIN3 MAX3]); ylim([MIN2 MAX2]);
279
280 subplot(3,3,7), plot(score(:,1), score(:,3), '.r')
281 xlim([MIN1 MAX1]); ylim([MIN3 MAX3]);
282 xlabel('PCA1');ylabel('PCA3');
283
284 subplot(3,3,8), plot(score(:,2), score(:,3), '.r')
285 xlim([MIN2 MAX2]); ylim([MIN3 MAX3]);
286 xlabel('PCA2');
287
288 subplot(3,3,9), plot(score(:,3), score(:,3), '.r')
289 xlim([MIN3 MAX3]); ylim([MIN3 MAX3]);
290 xlabel('PCA3');
291
292 %save([DIREC NOME '-dados-MFCC'] , 'coeff', 'score', 'latent',
      'tsquared', 'explained', 'mu');
293
294 figure;
295 for h = 1:length(list_population)
296     cor = rem(h,15);
297
298     if cor==1
```

```
299     pontos = [1 0 0];
300 elseif cor ==2
301     pontos = [0 0 1];
302 elseif cor ==3
303     pontos = [0 1 0];
304 elseif cor ==4
305     pontos = [0 0 0];
306 elseif cor ==5
307     pontos = [1 0 1];
308 elseif cor ==6
309     pontos = [1 1 0];
310 elseif cor ==7
311     pontos = [0 1 1];
312 elseif cor ==8
313     pontos = [0.5 0.5 1];
314 elseif cor ==9
315     pontos = [0.5 1 0.5];
316 elseif cor ==10
317     pontos = [0 0.5 0];
318 elseif cor ==11
319     pontos = [0.5 0 0.5];
320 elseif cor ==12
321     pontos = [0.5 0.5 0];
322 elseif cor ==13
323     pontos = [0.8 1 0.2];
324 elseif cor ==14
325     pontos = [0.2 0.6 0.7];
326 elseif cor ==0
327     pontos = [0.3 0.9 0.1];
328 end
329
330 plot3(score(categorias{h,2}+1:categorias{h+1,2},1), ...
```

```

331         score(categorias{h,2}+1:categorias{h+1,2},2), ...
332         score(categorias{h,2}+1:categorias{h+1,2},3),...
333         'Marker', '.', 'LineStyle', 'none', 'Color', pontos);
334     hold on;
335 end
336 grid on;
337 %title('SEM SUVREL');
338 xlabel('PCA 1');
339 ylabel('PCA 2');
340 zlabel('PCA 3');
341 % legend('subject A', 'subject B', 'subject C', 'subject D',
          'subject E', ...
342 %       'subject F', 'subject G', 'subject H', 'subject I', '
          subject J', ...
343 %       'subject K', 'subject L', 'subject M', 'subject N', '
          subject O');
344 hold off;
345
346 figure;
347 for h = 1:length(list_population)
348     raiosX(h,:) = std(score(categorias{h,2}+1:categorias{h
          +1,2},1:3));
349     centrosX(h,:) = mean(score(categorias{h,2}+1:categorias{h
          +1,2},1:3));
350 end
351
352 [x, y, z] = sphere;
353 for h = 1:length(list_population)
354     cor = rem(h,15);
355
356     if cor==1
357         superf = [1 0 0];

```

```
358     elseif cor ==2
359         superf = [0 0 1];
360     elseif cor ==3
361         superf = [0 1 0];
362     elseif cor ==4
363         superf = [0 0 0];
364     elseif cor ==5
365         superf = [1 0 1];
366     elseif cor ==6
367         superf = [1 1 0];
368     elseif cor ==7
369         superf = [0 1 1];
370     elseif cor ==8
371         superf = [0.5 0.5 1];
372     elseif cor ==9
373         superf = [0.5 1 0.5];
374     elseif cor ==10
375         superf = [0 0.5 0];
376     elseif cor ==11
377         superf = [0.5 0 0.5];
378     elseif cor ==12
379         superf = [0.5 0.5 0];
380     elseif cor ==13
381         superf = [0.8 1 0.2];
382     elseif cor ==14
383         superf = [0.2 0.6 0.7];
384     elseif cor ==0
385         superf = [0.3 0.9 0.1];
386     end
387
388     surf(x*(raiosX(h, 1)./PLOTFACTOR)+centrosX(h, 1), y*(
        raiosX(h, 2)./PLOTFACTOR)+centrosX(h, 2), z*(raiosX(h,
```

```

        3) ./PLOTFACTOR)+centrosX(h, 3), 'FaceAlpha',0.2,'
        LineStyle','none', 'FaceColor', superf );
389     hold on;
390 end
391
392 grid on;
393 title('SEM SUVREL');
394 xlabel('PCA 1');
395 ylabel('PCA 2');
396 zlabel('PCA 3');
397
398 % legend('subject A', 'subject B', 'subject C', 'subject D',
        'subject E', ...
399 %       'subject F', 'subject G', 'subject H', 'subject I', '
        subject J', ...
400 %       'subject K', 'subject L', 'subject M', 'subject N', '
        subject O');
401
402 hold off;
403
404 figure;
405 hold on;
406 t=-pi:0.01:pi;
407 for h = 1:length(list_population)
408     x0 = centrosX(h,1);
409     y0 = centrosX(h,2);
410     rx = (raiosX(h,1) ./PLOTFACTOR);
411     ry = (raiosX(h,2) ./PLOTFACTOR);
412     x = x0 + rx*cos(t);
413     y = y0 + ry*sin(t);
414
415     cor = rem(h,15);

```

```
416
417     if cor==1
418         pontos = [1 0 0];
419     elseif cor ==2
420         pontos = [0 0 1];
421     elseif cor ==3
422         pontos = [0 1 0];
423     elseif cor ==4
424         pontos = [0 0 0];
425     elseif cor ==5
426         pontos = [1 0 1];
427     elseif cor ==6
428         pontos = [1 1 0];
429     elseif cor ==7
430         pontos = [0 1 1];
431     elseif cor ==8
432         pontos = [0.5 0.5 1];
433     elseif cor ==9
434         pontos = [0.5 1 0.5];
435     elseif cor ==10
436         pontos = [0 0.5 0];
437     elseif cor ==11
438         pontos = [0.5 0 0.5];
439     elseif cor ==12
440         pontos = [0.5 0.5 0];
441     elseif cor ==13
442         pontos = [0.8 1 0.2];
443     elseif cor ==14
444         pontos = [0.2 0.6 0.7];
445     elseif cor ==0
446         pontos = [0.3 0.9 0.1];
447     end
```

```
448
449     plot(x,y,'Color', pontos);
450 end
451 hold off;
452
453 ab_vol = zeros(length(list_population), 1);
454 shared_volume = zeros(length(list_population));
455
456 for i=1:length(list_population)
457     for j=i:length(list_population)
458
459         [abv1, abv2, sv12, sv21] = ellipsoidVolumeComparison
460             (...
461                 centrosX(i,:), centrosX(j,:), raiosX(i,:), raiosX
462                     (j,:), 5e5);
463
464         ab_vol(i, 1) = abv1;
465
466         if i==j
467             shared_volume(i, j) = sv12;
468         else
469             shared_volume(i, j) = sv12;
470             shared_volume(j, i) = sv21;
471         end
472     end
473 end
474
475 figure1 = figure;
476 colormap('jet');
477 axes1 = axes('Parent',figure1);
478 hold(axes1,'on');
479 image(shared_volume,'Parent',axes1,'CDataMapping','scaled');
```

```
478 xlim(axes1,[0.5 length(list_population)+0.5]);
479 ylim(axes1,[0.5 length(list_population)+0.5]);
480 box(axes1,'on');
481 axis(axes1,'ij');
482 set(axes1,'Layer','top');
483 colorbar('peer',axes1);
484
485
486 %%
487
488 %WITH SUVREL
489
490 MULTIPLE_SAMPLES = 1; %For the case when it is known that n
    samples are from the same person
491 %DEFAULT VALUE = 1
492
493 %%%%%%%%%%
494 tic;
495 catg = 1;
496
497 CATEG = [];
498
499 for d = 1:categorias{end,2}
500     if d <= categorias{(catg*MULTIPLE_SAMPLES)+1,2}
501         CATEG(d,1) = catg;
502     end
503     if d == categorias{(catg*MULTIPLE_SAMPLES)+1,2}
504         catg = catg+1;
505     end
506 end
507
508 dadosSUVREL = [CATEG dadosParaPCA];
```

```
509
510 g = SUVREL(dadosSUVREL);
511 POTENCIA = 1;
512 dadosSUVRELParaPCA = dadosParaPCA * (g.^POTENCIA);
513
514 [coeff2, score2, latent2, tsquared2, explained2, mu2] = pca(
    dadosSUVRELParaPCA);
515 toc;
516
517 % figure;
518 % plot(1:1:length(latent2),latent2, '.b');
519
520 figure;
521 %limits:
522 MIN1 = min([score2(:,1); score2(:,2); score2(:,3)]);
523 MIN2 = MIN1;
524 MIN3 = MIN1;
525 MAX1 = max([score2(:,1); score2(:,2); score2(:,3)]);
526 MAX2 = MAX1;
527 MAX3 = MAX1;
528
529 subplot(3,3,1), plot(score2(:,1), score2(:,1), '.r')
530 xlim([MIN1 MAX1]); ylim([MIN1 MAX1]);
531 ylabel('PCA1');
532
533 subplot(3,3,2), plot(score2(:,2), score2(:,1), '.r')
534 xlim([MIN2 MAX2]); ylim([MIN1 MAX1]);
535
536 subplot(3,3,3), plot(score2(:,3), score2(:,1), '.r')
537 xlim([MIN3 MAX3]); ylim([MIN1 MAX1]);
538
539 subplot(3,3,4), plot(score2(:,1), score2(:,2), '.r')
```

```
540 xlim([MIN1 MAX1]); ylim([MIN2 MAX2]);
541 ylabel('PCA2');
542
543 subplot(3,3,5), plot(score2(:,2), score2(:,2), '.r')
544 xlim([MIN2 MAX2]); ylim([MIN2 MAX2]);
545
546 subplot(3,3,6), plot(score2(:,3), score2(:,2), '.r')
547 xlim([MIN3 MAX3]); ylim([MIN2 MAX2]);
548
549 subplot(3,3,7), plot(score2(:,1), score2(:,3), '.r')
550 xlim([MIN1 MAX1]); ylim([MIN3 MAX3]);
551 xlabel('PCA1');ylabel('PCA3');
552
553 subplot(3,3,8), plot(score2(:,2), score2(:,3), '.r')
554 xlim([MIN2 MAX2]); ylim([MIN3 MAX3]);
555 xlabel('PCA2');
556
557 subplot(3,3,9), plot(score2(:,3), score2(:,3), '.r')
558 xlim([MIN3 MAX3]); ylim([MIN3 MAX3]);
559 xlabel('PCA3');
560
561 figure;
562 for h = 1:length(list_population)
563     cor = rem(h,15);
564
565     if cor==1
566         pontos = [1 0 0];
567     elseif cor ==2
568         pontos = [0 0 1];
569     elseif cor ==3
570         pontos = [0 1 0];
571     elseif cor ==4
```

```
572     pontos = [0 0 0];
573 elseif cor ==5
574     pontos = [1 0 1];
575 elseif cor ==6
576     pontos = [1 1 0];
577 elseif cor ==7
578     pontos = [0 1 1];
579 elseif cor ==8
580     pontos = [0.5 0.5 1];
581 elseif cor ==9
582     pontos = [0.5 1 0.5];
583 elseif cor ==10
584     pontos = [0 0.5 0];
585 elseif cor ==11
586     pontos = [0.5 0 0.5];
587 elseif cor ==12
588     pontos = [0.5 0.5 0];
589 elseif cor ==13
590     pontos = [0.8 1 0.2];
591 elseif cor ==14
592     pontos = [0.2 0.6 0.7];
593 elseif cor ==0
594     pontos = [0.3 0.9 0.1];
595 end
596
597 plot3(score2(categorias{h,2}+1:categorias{h+1,2},1), ...
598     score2(categorias{h,2}+1:categorias{h+1,2},2), ...
599     score2(categorias{h,2}+1:categorias{h+1,2},3),...
600     'Marker', '.', 'LineStyle', 'none', 'Color', pontos);
601
602 hold on;
603 end
```

```

604 grid on;
605 title('COM SUVREL');
606 xlabel('PCA 1');
607 ylabel('PCA 2');
608 zlabel('PCA 3');
609 % legend('subject A', 'subject B', 'subject C', 'subject D',
        'subject E', ...
610 %       'subject F', 'subject G', 'subject H', 'subject I', '
        subject J', ...
611 %       'subject K', 'subject L', 'subject M', 'subject N', '
        subject O');
612
613 figure;
614 for h = 1:length(list_population)
615     raios(h,:) = std(score2(categorias{h,2}+1:categorias{h
        +1,2},1:3));
616     centros(h,:) = mean(score2(categorias{h,2}+1:categorias{h
        +1,2},1:3));
617 end
618
619 [x, y, z] = sphere;
620 for h = 1:length(list_population)
621     cor = rem(h,15);
622
623     if cor==1
624         superf = [1 0 0];
625     elseif cor ==2
626         superf = [0 0 1];
627     elseif cor ==3
628         superf = [0 1 0];
629     elseif cor ==4
630         superf = [0 0 0];

```

```
631     elseif cor ==5
632         superf = [1 0 1];
633     elseif cor ==6
634         superf = [1 1 0];
635     elseif cor ==7
636         superf = [0 1 1];
637     elseif cor ==8
638         superf = [0.5 0.5 1];
639     elseif cor ==9
640         superf = [0.5 1 0.5];
641     elseif cor ==10
642         superf = [0 0.5 0];
643     elseif cor ==11
644         superf = [0.5 0 0.5];
645     elseif cor ==12
646         superf = [0.5 0.5 0];
647     elseif cor ==13
648         superf = [0.8 1 0.2];
649     elseif cor ==14
650         superf = [0.2 0.6 0.7];
651     elseif cor ==0
652         superf = [0.3 0.9 0.1];
653     end
654
655     surf(x*(raios(h, 1)./PLOTFACTOR)+centros(h, 1), y*(raios(
        h, 2)./PLOTFACTOR)+centros(h, 2), z*(raios(h, 3)./
        PLOTFACTOR)+centros(h, 3), 'FaceAlpha',0.2,'LineStyle'
        , 'none', 'FaceColor', superf );
656     hold on;
657 end
658 grid on;
659
```

```
660 %title('COM SUVREL');
661 xlabel('PCA 1');
662 ylabel('PCA 2');
663 zlabel('PCA 3');
664 % legend('subject A', 'subject B', 'subject C', 'subject D',
        'subject E', ...
665 %       'subject F', 'subject G', 'subject H', 'subject I', '
        subject J', ...
666 %       'subject K', 'subject L', 'subject M', 'subject N', '
        subject O');

667
668 hold off;
669
670 figure;
671 hold on;
672 t=-pi:0.01:pi;
673 for h = 1:length(list_population)
674     x0 = centros(h,1);
675     y0 = centros(h,2);
676     rx = (raios(h,1)./PLOTFACTOR);
677     ry = (raios(h,2)./PLOTFACTOR);
678     x = x0 + rx*cos(t);
679     y = y0 + ry*sin(t);
680     plot(x,y);
681 end
682 hold off;
683
684 teste_g = sum(g(:).^2)
685
686 ab_volSUV = zeros(length(list_population), 1);
687 shared_volumeSUV = zeros(length(list_population));
688
```

```

689 for i=1:length(list_population)
690     for j=i:length(list_population)
691
692         [abv1SUV, abv2SUV, sv12SUV, sv21SUV] =
                ellipsoidVolumeComparison(...
693             centros(i,:), centros(j,:), raios(i,:), raios(j
                ,:), 5e5);
694
695         ab_volSUV(i, 1) = abv1SUV;
696
697         if i==j
698             shared_volumeSUV(i, j) = sv12SUV;
699         else
700             shared_volumeSUV(i, j) = sv12SUV;
701             shared_volumeSUV(j, i) = sv21SUV;
702         end
703     end
704 end
705
706 figure1 = figure;
707 colormap('jet');
708 axes1 = axes('Parent',figure1);
709 hold(axes1, 'on');
710 image(shared_volumeSUV, 'Parent', axes1, 'CDataMapping', 'scaled'
        );
711 xlim(axes1, [0.5 length(list_population)+0.5]);
712 ylim(axes1, [0.5 length(list_population)+0.5]);
713 box(axes1, 'on');
714 axis(axes1, 'ij');
715 set(axes1, 'Layer', 'top');
716 colorbar('peer', axes1);
717

```

```

718 sum(sum(shared_volumeSUV))/sum(sum(shared_volume))
719
720 %clearvars figure1;
721 save([DIRECTORY_MFCC 'allvariables.mat']);
722
723 %%
724
725 DIFFERENT_SPEAKER_HYP = [];
726
727 for i=1:size(shared_volume,2)
728     if i==1
729         DIFFERENT_SPEAKER_HYP(i,1) = mean(shared_volume(i,(i
730             +1):end));
731         DIFFERENT_SPEAKER_HYP(i,2) = std(shared_volume(i,(i
732             +1):end));
733     elseif i==size(shared_volume,2)
734         DIFFERENT_SPEAKER_HYP(i,1) = mean(shared_volume(i,1:
735             end-1));
736         DIFFERENT_SPEAKER_HYP(i,2) = std(shared_volume(i,1:
737             end-1));
738     else
739         DIFFERENT_SPEAKER_HYP(i,1) = mean([ shared_volume(i
740             ,1:(i-1)) shared_volume(i,(i+1):end) ]);
741         DIFFERENT_SPEAKER_HYP(i,2) = std([ shared_volume(i
742             ,1:(i-1)) shared_volume(i,(i+1):end) ]);
743     end
744 end
745
746 for i=1:size(shared_volumeSUV,2)
747     if i==1
748         DIFFERENT_SPEAKER_HYP_SUVREL(i,1) = mean(
749             shared_volumeSUV(i,(i+1):end));

```

```
743         DIFFERENT_SPEAKER_HYP_SUVREL(i,2) = std(  
            shared_volumeSUV(i,(i+1):end));  
744     elseif i==size(shared_volume,2)  
745         DIFFERENT_SPEAKER_HYP_SUVREL(i,1) = mean(  
            shared_volumeSUV(i,1:end-1));  
746         DIFFERENT_SPEAKER_HYP_SUVREL(i,2) = std(  
            shared_volumeSUV(i,1:end-1));  
747     else  
748         DIFFERENT_SPEAKER_HYP_SUVREL(i,1) = mean([  
            shared_volumeSUV(i,1:(i-1)) shared_volumeSUV(i,(i  
            +1):end) ]);  
749         DIFFERENT_SPEAKER_HYP_SUVREL(i,2) = std([  
            shared_volumeSUV(i,1:(i-1)) shared_volumeSUV(i,(i  
            +1):end) ]);  
750     end  
751 end  
752  
753 save([DIRECTORY_MFCC 'DIFFERENT_SPEAKER_HYP'], '  
        DIFFERENT_SPEAKER_HYP');  
754 save([DIRECTORY_MFCC 'DIFFERENT_SPEAKER_HYP_SUVREL'], '  
        DIFFERENT_SPEAKER_HYP_SUVREL');  
755  
756  
757 %% LR CALCULATION  
758  
759 DIRECTORY_SAVE_SAME_SPEAKER_HYP = 'F:\SUVREL\FINAL-female-  
        SNR25\';  
760 load([DIRECTORY_SAVE_SAME_SPEAKER_HYP 'SAME_SPEAKER_HYP.mat'  
        ]);  
761 load([DIRECTORY_SAVE_SAME_SPEAKER_HYP 'EVIDENCE']);  
762  
763 load([DIRECTORY_MFCC 'DIFFERENT_SPEAKER_HYP']);
```

```

764 load([DIRECTORY_MFCC 'DIFFERENT_SPEAKER_HYP_SUVREL']);
765
766 mkdir([DIRECTORY_SAVE 'plots']);
767
768 PLOT_LR = 1; %change to 1 to get all LR plots
769 x = linspace(0,1,1e2);
770 LR = [];
771
772 for i=1:size(SAME_SPEAKER_HYP,1)
773     HP = normpdf(EVIDENCE(i,1), SAME_SPEAKER_HYP(i,1),
774                 SAME_SPEAKER_HYP(i,2));
775     HD_without_suvrel = normpdf(EVIDENCE(i,1),
776                                DIFFERENT_SPEAKER_HYP(i,1), DIFFERENT_SPEAKER_HYP(i,2)
777                                );
778     HD_with_suvrel = normpdf(EVIDENCE(i,1),
779                             DIFFERENT_SPEAKER_HYP_SUVREL(i,1),
780                             DIFFERENT_SPEAKER_HYP_SUVREL(i,2));
781
782     LR(i,1) = HP/HD_without_suvrel;
783     LR(i,2) = HP/HD_with_suvrel;
784     LR(i,3) = ((LR(i,2)/LR(i,1))-1)*100;
785
786     if PLOT_LR == 1
787         figure;
788         hold on;
789         set(gcf,'position',[10 10 550 400])
790
791         y1 = normpdf(x,SAME_SPEAKER_HYP(i,1),
792                     SAME_SPEAKER_HYP(i,2));
793         y2 = normpdf(x,DIFFERENT_SPEAKER_HYP(i,1),
794                     DIFFERENT_SPEAKER_HYP(i,2));
795         plot(x,y1,'ob',x,y2,'.r');

```

```

789     y = ylim;
790     plot([EVIDENCE(i,1) EVIDENCE(i,1)],[y(1) y(2)], '-g')
791
792     title({'Indivduo: ' num2str(i) ' / UBM: ' num2str(
        NUMBER_OF_SUBJECTS) ' ' GENDER ' / freq = '
        num2str(RESAMPLEAUDIO/1000) 'kHz / SNR = ' num2str
        (SNR)]; ' (sem SUVREL)'},'FontSize',14);
793     xlabel('Intersec de volumes das elipses (%)','
        FontSize',16);
794     ylabel('Frequencia ','FontSize',16);
795     legend({'H_{P}', 'H_{D}'}, ['E = ' num2str(EVIDENCE(i
        ,1))]}, 'FontSize',15, ...
796         'Position',[0.200454548109662 0.441250003576279
        0.252727267417041 0.252499992847443]);
797
798     dim = [0.2 0.729000001937151 0.252727267417041
        0.0899999980628491];
799     str = ['LR = ' num2str(LR(i,1))];
800     annotation('textbox',dim,'String',str,'FitBoxToText',
        'on', 'FontSize',15);
801
802     box('on');
803     hold off;
804     saveas(gcf, [DIRECTORY_SAVE 'plots\' ['semSUVREL_' '
        ind' num2str(i) '_UBM' num2str(NUMBER_OF_SUBJECTS)
        GENDER '_freq' num2str(RESAMPLEAUDIO/1000) '
        kHz_SNR' num2str(SNR) ] ], 'eps');
805     saveas(gcf, [DIRECTORY_SAVE 'plots\' ['semSUVREL_' '
        ind' num2str(i) '_UBM' num2str(NUMBER_OF_SUBJECTS)
        GENDER '_freq' num2str(RESAMPLEAUDIO/1000) '
        kHz_SNR' num2str(SNR) ] ], 'jpg');
806     saveas(gcf, [DIRECTORY_SAVE 'plots\' ['semSUVREL_' '

```

```

    ind' num2str(i) '_UBM' num2str(NUMBER_OF_SUBJECTS)
      GENDER '_freq' num2str(RESAMPLEAUDIO/1000) '
      kHz_SNR' num2str(SNR) ] ], 'fig');
807
808 figure;
809 hold on;
810 set(gcf,'position',[10 10 550 400])
811
812 y1 = normpdf(x,SAME_SPEAKER_HYP(i,1),
      SAME_SPEAKER_HYP(i,2));
813 y2 = normpdf(x,DIFFERENT_SPEAKER_HYP_SUVREL(i,1),
      DIFFERENT_SPEAKER_HYP_SUVREL(i,2));
814 plot(x,y1,'ob',x,y2,'.r');
815 y = ylim;
816 plot([EVIDENCE(i,1) EVIDENCE(i,1)],[y(1) y(2)], '-g')
817
818 title({'Indivduo: ' num2str(i) ' / UBM: ' num2str(
      NUMBER_OF_SUBJECTS) ' ' GENDER ' / freq = '
      num2str(RESAMPLEAUDIO/1000) 'kHz / SNR = ' num2str
      (SNR)]; ' (com SUVREL)'},'FontSize',14);
819 xlabel('Intersec de volumes das elipses (%)','
      FontSize',16);
820 ylabel('Frequencia','FontSize',16);
821 legend({'H_{P}','H_{D}', ['E = ' num2str(EVIDENCE(i
      ,1))]}}, 'FontSize',15, ...
822       'Position',[0.200454548109662 0.441250003576279
      0.252727267417041 0.252499992847443]);
823
824 dim = [0.2 0.729000001937151 0.252727267417041
      0.0899999980628491];
825 str = ['LR = ' num2str(LR(i,2))];
826 annotation('textbox',dim,'String',str,'FitBoxToText',

```

```

        'on', 'FontSize',15);
827
828     box('on');
829     hold off;
830     saveas(gcf, [DIRECTORY_SAVE 'plots\' ['comSUVREL_' '
        ind' num2str(i) '_UBM' num2str(NUMBER_OF_SUBJECTS)
        GENDER '_freq' num2str(RESAMPLEAUDIO/1000) '
        kHz_SNR' num2str(SNR) ] ], 'eps');
831     saveas(gcf, [DIRECTORY_SAVE 'plots\' ['comSUVREL_' '
        ind' num2str(i) '_UBM' num2str(NUMBER_OF_SUBJECTS)
        GENDER '_freq' num2str(RESAMPLEAUDIO/1000) '
        kHz_SNR' num2str(SNR) ] ], 'jpg');
832     saveas(gcf, [DIRECTORY_SAVE 'plots\' ['comSUVREL_' '
        ind' num2str(i) '_UBM' num2str(NUMBER_OF_SUBJECTS)
        GENDER '_freq' num2str(RESAMPLEAUDIO/1000) '
        kHz_SNR' num2str(SNR) ] ], 'fig');
833     end
834 end
835
836 close all;
837 %RESULTS
838 % [subject EVIDENCE LR_without_SUVREL LR_with_SUVREL
        LR_variation%]
839 results = [ [1:i]' EVIDENCE LR];
840
841 save([DIRECTORY_SAVE 'results'], 'results');

```

B.3. Função “SUVREL.m”

```

1 function [g] = SUVREL(dadosSUVREL)
2 GAMMA = 2;
3 K = max(dadosSUVREL(:,1));

```

```

4 medias = zeros(max(dadosSUVREL(:,1)), size(dadosSUVREL,2)-1);
5
6 for u = 1:(size(dadosSUVREL,2)-1)
7     for a = 1:K
8         medias(a, u) = mean(dadosSUVREL(find(dadosSUVREL(:,1)
9             ==a),u+1));
10    end
11
12 covariancias_m = zeros((size(dadosSUVREL,2)-1), (size(
13     dadosSUVREL,2)-1));
14 for u = 1:(size(dadosSUVREL,2)-1)
15     for v = 1:(size(dadosSUVREL,2)-1)
16         soma_medias = 0;
17         for a = 1:K
18             soma_medias = soma_medias + medias(a, u) * medias
19                 (a, v);
20         end
21         covariancias_m(u, v) = (1/K)*soma_medias - ( ((1/K)*
22             sum(medias(:,u))) * ((1/K)*sum(medias(:,v))) );
23     end
24 end
25 covariancias_x = zeros((size(dadosSUVREL,2)-1), (size(
26     dadosSUVREL,2)-1),max(dadosSUVREL(:,1)));
27
28 for a = 1:K
29     for u = 1:(size(dadosSUVREL,2)-1)
30         for v = 1:(size(dadosSUVREL,2)-1)
31             w = cov(dadosSUVREL(find(dadosSUVREL(:,1)==a),u
32                 +1),dadosSUVREL(find(dadosSUVREL(:,1)==a),v+1)
33                 );

```

```

29         covariancias_x(u, v, a) = sum( w(:) );
30     end
31 end
32 end
33
34 e_in = 2*sum(covariancias_x,3);
35 e_out = ((K-1)*sum(covariancias_x,3)) + ((K^2) .*
        covariancias_m);
36 eps = e_in + (GAMMA*e_out);
37 eps_sq = eps.^2;
38 norm = sqrt(sum(eps_sq(:)));
39 for u = 1:(size(dadosSUVREL,2)-1)
40     for v = 1:(size(dadosSUVREL,2)-1)
41         g(u,v) = -eps(u,v)/norm;
42     end
43 end

```

B.4. Função “ellipsoidVolumeComparison.m”

```

1 function [absolute_volume1, absolute_volume2, shared_volume12,
        shared_volume21] = ellipsoidVolumeComparison(center1,
        center2, radius1, radius2, N, varargin)
2 %N=5000000;
3
4 %center1 = centrosX(1,:);
5 %center2 = centrosX(2,:);
6
7 % radius1 = raiosX(1,:);
8 % radius2 = raiosX(2,:);
9
10 MINx = min([center1(1);center2(1)]);
11 MAXx = max([center1(1);center2(1)]);

```

```

12 MINy = min([center1(2);center2(2)]);
13 MAXy = max([center1(2);center2(2)]);
14 MINz = min([center1(3);center2(3)]);
15 MAXz = max([center1(3);center2(3)]);
16
17 STDx = max([radius1(1);radius2(1)]);
18 STDy = max([radius1(1);radius2(1)]);
19 STDz = max([radius1(1);radius2(1)]);
20
21 limits = [MINx-STDx MAXx+STDx;
22           MINy-STDy MAXy+STDy;
23           MINz-STDz MAXz+STDz];
24
25 x = limits(1,1) + (limits(1,2) - limits(1,1))*rand(N,1);
26 y = limits(2,1) + (limits(2,2) - limits(2,1))*rand(N,1);
27 z = limits(3,1) + (limits(3,2) - limits(3,1))*rand(N,1);
28
29 x1 = x - center1(1);
30 y1 = y - center1(2);
31 z1 = z - center1(3);
32
33 x2 = x - center2(1);
34 y2 = y - center2(2);
35 z2 = z - center2(3);
36
37 r1 = (x1/radius1(1)).^2 + (y1/radius1(2)).^2 + (z1/radius1(3)
38       ).^2;
39 r2 = (x2/radius2(1)).^2 + (y2/radius2(2)).^2 + (z2/radius2(3)
40       ).^2;
41 %r2 = sqrt(x2.^2 + y2.^2 + z2.^2);
42
43 m12 = 0; %Number of points inside spheres 1 and 2

```

```
42 m1 = 0;    %Number of points inside sphere1
43 m2 = 0;    %Number of points inside sphere2
44 n_in = 1;
45 n_out1 = 1;
46 n_out2 = 1;
47
48 %in = [];
49 in = zeros(1,N);
50 out1 = zeros(1,N);
51 out2 = zeros(1,N);
52
53 for i = 1:N
54     if r1(i)<=1 && r2(i)<=1
55         m12 = m12 + 1;
56         in(n_in) = i;
57         n_in = n_in + 1;
58     elseif r1(i)<=1
59         m1 = m1 + 1;
60         out1(n_out1) = i;
61         n_out1 = n_out1 + 1;
62     elseif r2(i)<=1
63         m2 = m2 + 1;
64         out2(n_out2) = i;
65         n_out2 = n_out2 + 1;
66     end
67 end
68
69 if nargin == 6
70     figure;
71     hold on;
72     axis square;
73     in = in(1:find(in==max(in)));
```

```

74     out1 = out1(1:find(out1==max(out1)));
75     out2 = out2(1:find(out2==max(out2)));
76
77     if sum((center1 == center2))+ sum((radius1 == radius2))
       == 6
78         plot3(x(in),y(in),z(in),'Marker', '.', 'LineStyle', '
           none', 'Color', 'b');
79     else
80         plot3(x(in),y(in),z(in),'Marker', '.', 'LineStyle', '
           none', 'Color', 'b');
81         plot3(x(out1),y(out1),z(out1),'Marker', '.', '
           LineStyle', 'none', 'Color', 'y');
82         plot3(x(out2),y(out2),z(out2),'Marker', '.', '
           LineStyle', 'none', 'Color', 'c');
83     end
84     grid on;
85     xlabel('PCA 1');
86     ylabel('PCA 2');
87     zlabel('PCA 3');
88     hold off;
89 end
90
91 cube_volume = (limits(1,2) - limits(1,1))*(limits(2,2) -
    limits(2,1))*...
92     (limits(3,2) - limits(3,1));
93 absolute_volume1 = ((m1 + m12)/N)*cube_volume;
94 absolute_volume2 = ((m2 + m12)/N)*cube_volume;
95 shared_volume12 = m12/(m1 + m12);
96 shared_volume21 = m12/(m2 + m12);

```

B.5. Função “ellipsoidVolumeComparisonErrorTest.m”

```

1 %%
2 N =100;
3 mSV12STDx = zeros(N,1);
4
5 for j = 1:100
6     ABV1 = zeros(N,1);
7     ABV2 = zeros(N,1);
8     SV12 = zeros(N,1);
9     SV21 = zeros(N,1);
10
11     for i=1:N
12         [abv1, abv2, sv12, sv21] = ellipsoidVolumeComparison
13             (...
14                 [0.5 0.5 0.5], [0.9 0.5 0.5], [0.5 0.5 0.5], [0.9
15                     0.5 0.5], 5e5);
16         ABV1(i,1) = abv1;
17         ABV2(i,1) = abv2;
18         SV12(i,1) = sv12;
19         SV21(i,1) = sv21;
20         % [abv1, abv2, sv12, sv21] =
21             ellipsoidVolumeComparison(...
22                 %             centrosX(i,:), centrosX(j,:), raiosX(i
23                     ,:), raiosX(j,:), 5e5);
24     end
25
26     m1 = mean(ABV1);
27     m1std = std(ABV1);
28     mSV12 = mean(SV12);
29     mSV12std = std(SV12);
30     mSV12STDx(j,1) = mSV12std/mSV12;
31 end

```

```

29 mean(mSV12STDx)
30 std(mSV12STDx)
31 std(mSV12STDx)/mean(mSV12STDx)

```

B.6. Função “melTriangularFilterBank.m”

```

1 function[H, hi, mi] = melTriangularFilterBank(num_filt,
        samplelength, lowerfreq, samplerate, varargin)
2 %[H, hi, mi] = melTriangularFilterBank(num_filt, samplelength
        , lowerfreq, samplerate)
3 %     returns: 1) a MxN vector H where N is the number of
        the filter ans M are its
4 %     values, hi are the frequencies of each filter bank and
        mi are hi
5 %     correspondants in mel-scale.
6 %     lowerfreq must be informed. upperfreq equal samplerate
        /2.
7 %
8 %[H, hi, mi] = melTriangularFilterBank(num_filt, samplelength
        , lowerfreq, samplerate, 'plot')
9 %     returns: 1) a MxN vector H where N is the number of
        the filter ans M are its
10 %    values, hi are the frequencies of each filter bank and
        mi are hi
11 %    correspondants in mel-scale.
12 %    lowerfreq must be informed. upperfreq equal samplerate
        /2.
13 %    Also, returns an illustrative plot of the triangular
        filter-bank made.
14
15 % samplerate = 44100;
16 % num_filt = 10;

```

```

17
18 upperfreq = samplerate/2;
19 lowermel = freq2melscale(lowerfreq);
20 uppermel = freq2melscale(upperfreq);
21 mi = linspace(lowermel, uppermel, num_filt+2);
22 hi = melscale2freq(mi);
23 fi = floor((samplelength+1)*hi)/samplerate;
24 %fi = floor((nfft).*hi/samplerate);
25 x_fft_bin = linspace(1, samplelength/2+1, samplelength/2+1);
26 H = zeros(samplelength/2+1, num_filt);
27
28 for m = 2:1:(num_filt+1)
29     for i = 1:1:(samplelength/2+1)
30         if ( i >= fi(m-1) && i <= fi(m) )
31             H(i, m-1) = ( i - fi(m-1) ) / ( fi(m) - fi(m-1) );
32         elseif ( i >= fi(m) && i <= fi(m+1) )
33             H(i, m-1) = ( fi(m+1) - i ) / ( fi(m+1) - fi(m) );
34         end
35     end
36 end
37
38 %%%%%%%%%%%
39 if nargin == 5
40     H_plot = zeros(upperfreq, num_filt);
41     x_freq = linspace(1, upperfreq, upperfreq);
42     figure; box on; grid on;
43     hold on;
44     for m = 2:1:(num_filt+1)
45         for i = 1:1:(upperfreq)
46             if ( i >= hi(m-1) && i <= hi(m) )
47                 H_plot(i, m-1) = ( i - hi(m-1) ) / ( hi(m) - hi(
m-1) );

```

```

48         elseif ( i >= hi(m) && i <= hi(m+1) )
49             H_plot(i, m-1) = (hi(m+1) - i)/(hi(m+1) - hi(
                    m));
50         end
51     end
52 end
53 for filt = 1:num_filt
54     plot(x_freq, H_plot(:,filt))
55     ax = gca;
56     ax.XAxis.Exponent = 0;
57     ax.XAxis.TickLabelFormat = '%6.0f';
58 end
59 title([num2str(num_filt) '-channel Triangular Mel-Filter
        Bank'])
60 xlabel('frequency(Hz)')
61 ylabel('Filter Gain')
62 hold off;
63 end

```

B.7. Função “freq2melscale.m”

```

1 function [melscale] = freq2melscale (freq)
2 % melscale] = freq2melscale (freq)
3
4 melscale = 1125 * log(1 + freq/700);

```

B.8. Função “melscale2freq.m”

```

1 function [freq] = melscale2freq (melscale)
2 % melscale] = freq2melscale (freq)
3
4 freq = 700 * (exp(melscale/1125) - 1);

```