



DETECÇÃO DE DEFEITOS EM DUTOS RÍGIDOS POR EMISSÃO ACÚSTICA E REDES NEURAIIS

Marco Antonio Arias Suárez

Dissertação de Mestrado apresentado ao Programa de Pós-graduação em Engenharia Elétrica, COPPE, da Universidade Federal do Rio de Janeiro, como parte dos requisitos necessários à obtenção do título de Mestre em Engenharia Elétrica.

Orientador: Luiz Pereira Calôba

Rio de Janeiro
Dezembro de 2015

DETECÇÃO DE DEFEITOS EM DUTOS RÍGIDOS POR EMISSÃO ACÚSTICA E
REDES NEURAIIS

Marco Antonio Arias Suárez

DISSERTAÇÃO SUBMETIDA AO CORPO DOCENTE DO INSTITUTO ALBERTO
LUIZ COIMBRA DE PÓS-GRADUAÇÃO E PESQUISA DE ENGENHARIA
(COPPE) DA UNIVERSIDADE FEDERAL DO RIO DE JANEIRO COMO PARTE
DOS REQUISITOS NECESSÁRIOS PARA A OBTENÇÃO DO GRAU DE MESTRE
EM CIÊNCIAS EM ENGENHARIA ELÉTRICA.

Examinada por:

Prof. Jorge Lopes de Souza Leão, D.Sc.

Dr. Romeu Ricardo da Silva, D.Sc.

Dr. Sergio Damasceno Soares, D.Sc.

RIO DE JANEIRO, RJ - BRASIL

DEZEMBRO DE 2015

Arias, Marco Antonio

Detecção de Defeitos em Dutos Rígidos por Emissão Acústica e Redes Neurais / Arias, Marco Antonio. Rio de Janeiro: UFRJ/COPPE, 2015.

XV, 88 p. il.; 29,7cm

Orientador: Luiz Pereira Calôba

Dissertação (mestrado) – UFRJ / COPPE/ Programa de Engenharia Elétrica, 2015.

Referências Bibliográficas: p. 86-87

1. Classificação. 2. Emissão Acústica. I. Calôba, Luiz Pereira. II. Universidade Federal do Rio de Janeiro, COPPE, Programa de Engenharia Elétrica. III. Título.

Dedicatória

A minha mãe Isabel, que me ensinou a sonhar e a lutar para tornar esses sonhos em realidade, e por ser meu apoio em todo momento.

Ao meu pai Juan, que me ensinou a amar a engenharia.

Agradecimentos

Primeiramente a Deus, pelas forças que me deu nos momentos difíceis e por todos os presentes que me dá a cada dia, especialmente aqueles sutéis.

Ao meu orientador, pelos ensinamentos e pelo apoio em todo momento para o desenvolvimento deste trabalho.

À família Moustapha, que tornou-se minha família no Brasil e que me deram todo seu apoio para concluir esta etapa da minha vida.

Aos meus amigos pela ajuda nesta pesquisa, especialmente ao Jorge, ao Miguel e ao Leiz.

Resumo da Dissertação apresentada à COPPE/UFRJ como parte dos requisitos necessários para a obtenção do grau de Mestre em Ciências (M.Sc.)

DETECÇÃO DE DEFEITOS EM DUTOS RÍGIDOS POR EMISSÃO ACÚSTICA E REDES NEURAIAS

Marco Antonio Arias Suárez

Dezembro/2015

Orientador: Luiz Pereira Calôba

Programa: Engenharia Elétrica

Nesta dissertação desenvolvemos uma interface *SCADA* para a detecção de defeitos em dutos rígidos utilizando ensaio não destrutivo via emissão acústica. O sistema *SCADA* compreendeu o desenvolvimento de um sistema de aquisição de dados para os sensores de emissão acústica, utilizando uma placa da *National Instruments* e a plataforma *LabView* para a programação da interface. Este sistema foi utilizado durante um ensaio realizado no LAMEF na UFRGS, como parte de um projeto com o CENPES para monitoramento de defeitos em dutos rígidos. A informação adquirida foi analisada por um classificador não supervisionado, cuja finalidade era de identificar se o defeito estrutural do duto encontra-se no estado de propagação estável ou em propagação instável. Os resultados obtidos apresentaram índices de acerto em torno de 88%, sugerindo a viabilidade de uma implementação de um sistema em tempo real.

Abstract of Dissertation presented to COPPE/UFRJ as a partial fulfillment of the requirements for the degree of Master of Science (M.Sc.)

DETECTION OF DEFECTS OF RIGIDS PIPES BY ACOUSTIC EMISSION AND
NEURAL NETWORKS

Marco Antonio Arias Suárez

December/2015

Advisor: Luiz Pereira Calôba

Department: Electrical Engineering

In this work was developed a SCADA interface for defect detection on rigid pipes using non-destructive testing by acoustic emission. The SCADA system included the implementation of the data acquisition system for acoustic emission sensors, using a National Instruments' device and the LabView software to develop the interface. This system was used during a test at LAMEF laboratory of UFRGS, as a part of the CENPES' Project for defects monitoring in rigid pipelines. The acquired data were analyzed by a unsupervised classifier whose purpose was to identify if the structural defect was in stable or in unstable propagation. The results showed hit rates around 88%, suggesting the feasibility to implementing a real time system.

SUMÁRIO

1	Introdução.....	1
1.1	Contextualização.....	1
1.2	Problema/Motivação.....	2
1.3	Objetivos.....	4
1.4	Justificativa.....	4
2	Revisão Bibliográfica.....	5
2.1	Emissão Acústica.....	5
2.1.1	Propagação de Ondas Acústicas.....	6
2.1.2	Tipos de Sinais de Emissão Acústica.....	8
2.1.3	Parâmetros de EA.....	9
2.1.4	Processamento de Sinais de EA.....	11
2.2	Redes Neurais Artificiais.....	12
2.2.1	Aprendizado Supervisionado.....	13
2.2.2	Aprendizado Não Supervisionado.....	14
2.2.2.1	<i>K-Means</i>	14
2.3	Mecânica da fratura em dutos.....	15
3	Metodologia.....	19
3.1	Implementação do Sistema de Aquisição de Sinais de EA.....	19
3.1.1	Equipamentos.....	19
3.1.2	Software de Aquisição.....	21
3.1.3	Descrição do Ensaio.....	23
3.1.4	Definição das Classes de trinca.....	25
3.1.5	Pré – Processamento dos dados.....	26
3.2	Desenvolvimento do Classificador.....	26
3.2.1	Breve descrição do Ensaio 3.....	26

3.2.2	O Classificador	28
3.3	Análise dos grupos classificados	30
4	Resultados	32
4.1	Resultados da Implementação do Sistema de Aquisição de Sinais de EA	32
4.2	Classificação usando um conjunto de 36 dados.....	35
4.2.1	Caso 1: Classificação Árvore sem reclassificação	37
4.2.2	Caso 2: Classificação árvore sem reclassificação e inicializando os centroides.....	41
4.2.3	Caso 3: Classificação Árvore com reclassificação	44
4.2.4	Caso 4: Classificação Árvore com reclassificação e inicializando os centroides.....	46
4.3	Classificação usando um conjunto de 68 dados.....	48
4.3.1	Caso 5: Classificação Árvore sem reclassificação	50
4.3.2	Caso 6: Classificação árvore sem reclassificação e inicializando os centroides.....	54
4.3.3	Caso 7: Classificação Árvore com reclassificação	57
4.3.4	Caso 8: Classificação Árvore com reclassificação e inicializando os centroides.....	61
4.4	Validação Cruzada.....	63
4.4.1	Caso 1: Classificação com centroides aleatórios.....	64
4.4.2	Caso 2: Classificação com inicializando os centroides	71
4.5	Acurácia dos Classificadores desenvolvidos com 36 dados.....	78
4.6	Comparação de Resultados com outros trabalhos	82
5	Considerações.....	83
6	Conclusões.....	84
7	Trabalhos Futuros.....	85
8	Referência Bibliográfica.....	86
	APÊNDICE	88

ÍNDICE DE FIGURAS

Figura 1. Processo de Produção <i>Offshore</i> de Petróleo [8].....	1
Figura 2. Tubo rígido com costura. Fonte: Palestra de Instalações Submarinos.....	3
Figura 3. Características da Emissão Acústica [3].	5
Figura 4. Diagrama de Bloco do Sensor Integrado [6].....	6
Figura 5. Sinal Transitório [15].	8
Figura 6. Sinal contínuo [15].....	8
Figura 7. Parâmetros de EA [9].....	9
Figura 8. Diagrama de Blocos do Processamento de Sinais [9].....	12
Figura 9. Diagrama de Aprendizado Supervisionado [17].....	13
Figura 10. Diagrama de Aprendizado Não Supervisionado [17].	14
Figura 11. Algoritmo do Método <i>K-Means</i> [15].	15
Figura 12. Linha de Contorno em torno da Trinca [13].	16
Figura 13. Curva J esquemática para um material dúctil [13].....	16
Figura 14. Duto com trinca circunferencial [13].	18
Figura 15. Equipamentos de Aquisição: (a) Sensor R15I-AST, (b) Placa NI PCI-6250 e (c) Bloco conector SCB-68A.....	20
Figura 16. Circuito de conexão dos sensores com o sistema de aquisição.....	21
Figura 17. Interface do Sistema de Aquisição de Sinais de EA.	22
Figura 18. Interface Diadem.	22
Figura 19. Duto com sensores de EA montados.....	23
Figura 20. Distribuição dos sensores no duto.....	27
Figura 21. Curva de Pressão e Deformação vs tempo.....	28
Figura 22. Classificação com estrutura tipo árvore.	29
Figura 23. Classificação <i>K-Means</i> reclassificando o total de dados.....	30
Figura 24. Validação Cruzada.	31
Figura 25. Circuito de Alimentação do Sensor [5].....	33
Figura 26. Duto de 2m de comprimento instrumentado [18].	34
Figura 27. Forma de Onda PE.	35
Figura 28. Forma de Onda PI.	36
Figura 29. Entradas PE e PI.....	38
Figura 30. Classificação com 2 classes para o Caso 1.....	38
Figura 31. Classificação com 3 Classes para o Caso 1.....	39

Figura 32. Classificação com 4 Classes para o Caso 1.....	40
Figura 33. Classificação com 2 Classes para o Caso 2.....	41
Figura 34. Classificação com 3 Classes para o Caso 2.....	42
Figura 35. Classificação com 5 Classes para o Caso 2.....	43
Figura 36. Classificação com 3 Classes para o Caso 3.....	44
Figura 37. Classificação com 4 Classes para o Caso 3.....	45
Figura 38. Classificação com 3 Classes para o Caso 4.....	46
Figura 39. Classificação com 5 Classes para o Caso 4.....	47
Figura 40. Forma de Onda PE.	49
Figura 41. Forma de Onda PI.	49
Figura 42. Ondas PE e PI.	50
Figura 43. Classificação com duas Classes para o Caso 5.	51
Figura 44. Classificação com 4 Classes para o Caso 5.....	52
Figura 45. Classificação com 6 Classes para o Caso 5.....	53
Figura 46. Classificação com 2 Classes para o Caso 6.....	54
Figura 47. Classificação com 4 Classes para o Caso 6.....	55
Figura 48. Classificação com 6 Classes para o Caso 6.....	56
Figura 49. Classificação com 7 Classes para o Caso 6.....	57
Figura 50. Classificação com 2 Classes para o Caso 7.....	58
Figura 51. Classificação com 4 Classes para o Caso 7.....	59
Figura 52. Classificação com 6 Classes para o Caso 7.....	60
Figura 53. Classificação com 4 Classes para o Caso 8.....	61
Figura 54. Classificação com 6 Classes para o Caso 8.....	62
Figura 55. Ondas PE e PI das partições A e B.	64
Figura 56. Classificação com 2 classes para as partições A e B.....	65
Figura 57. Classificação com 4 Classes para as partições A e B.....	65
Figura 58. Classificação com 6 Classes para as partições A e B.....	66
Figura 59. Teste com a partição C.	66
Figura 60. Formas de Onda PE e PI das partições A e C.	67
Figura 61. Classificação com 6 Classes para as partições A e C.....	68
Figura 62. Teste com a partição B.....	68
Figura 63. Formas de onda PE e PI para as partições B e C.	69
Figura 64. Classificação com 6 Classes para as partições B e C.....	70
Figura 65. Teste com a partição A.....	70

Figura 66. Formas de Onda PE e PI das partições A e B.	72
Figura 67. Classificação com 6 Classes para as partições A e B.	72
Figura 68. Teste com a partição C.	73
Figura 69. Formas de Onda PE e PI para as partições A e C.	74
Figura 70. Classificação com 6 Classes para as partições A e C.	74
Figura 71. Teste com a partição B.	75
Figura 72. Formas de Onda PE e PI para as partições B e C.	76
Figura 73. Classificação com 6 Classes para as partições B e C.	76
Figura 74. Teste com a partição A.	77
Figura 75. Classificação com os centroides do Caso 1.	78
Figura 76. Classificação com os centroides do Caso 2.	79
Figura 77. Classificação com os centroides do Caso 3.	80
Figura 78. Classificação com os centroides do Caso 4.	81

ÍNDICE DE TABELAS

Tabela 1. Principais Parâmetros de EA [9].	10
Tabela 2. Fatores que influenciam a EA. [3].....	11
Tabela 3. Parâmetros de EA adquiridos pelo sistema PAC [3].	24
Tabela 4. Quantidade dos sensores EA sobre o duto.....	27
Tabela 5. Índice de Acerto para o Caso 1 com duas classes.	39
Tabela 6. Índice de Acerto para o Caso 1 com 3 classes.....	39
Tabela 7. Índice de Acerto para o Caso 1 com 4 Classes.....	40
Tabela 8. Índice de Acerto para o Caso 2 com 2 Classes.....	41
Tabela 9. Índice de Acerto para o Caso 2 com 3 Classes.....	42
Tabela 10. Índice de Acerto para o Caso 2 com 5 Classes.....	43
Tabela 11. Índice de Acerto para o Caso 3 com 3 Classes.....	44
Tabela 12. Índice de Acerto para o Caso 3 com 4 Classes.....	45
Tabela 13. Índice de Acerto para o Caso 4 com 3 Classes.....	46
Tabela 14. Índice de Acerto para o Caso 4 com 5 Classes.....	47
Tabela 15. Índice de acerto para o Caso 5 com 2 Classes.....	51
Tabela 16. Índice de Acerto para o Caso 5 com 4 Classes.....	52
Tabela 17. Índice de Acerto para o Caso 5 com 5 Classes.....	53
Tabela 18. Índice de Acerto para o Caso 6 com 2 Classes.....	54
Tabela 19. Índice de Acerto para o Caso 6 com 4 Classes.....	55
Tabela 20. Índice de Acerto para o Caso 6 com 6 Classes.....	56
Tabela 21. Índice de Acerto para o Caso 6 com 7 Classes.....	57
Tabela 22. Índice de Acerto para o Caso 7 com 2 Classes.....	58
Tabela 23. Índice de Acerto para o Caso 7 com 4 Classes.....	59
Tabela 24. Índice de Acerto para o Caso 7 com 6 Classes.....	60
Tabela 25. Índice de Acerto para o Caso 8 com 4 Classes.....	61
Tabela 26. Índice de Acerto para o Caso 8 com 6 Classes.....	62
Tabela 27. Índice de Acerto com a Partição C.	67
Tabela 28. Índice de Acerto da partição B.	69
Tabela 29. Índice de Acerto com a partição A.	71
Tabela 30. Eficiência de Predição.	71
Tabela 31. Índice de Acerto para a partição C.	73
Tabela 32. Índice de Acerto para a Partição B.	75

Tabela 33. Índice de Acerto com a partição A.	77
Tabela 34. Eficiência de Predição.	77
Tabela 35. Índice de Acerto com os centroides do Caso 1.	79
Tabela 36. Índice de Acerto com os centroides do Caso 2.	80
Tabela 37. Índice de Acerto com os centroides do Caso 3.	81
Tabela 38. Índice de Acerto com os centroides do Caso 4.	82

Nomenclatura

- SCADA* - *Supervisory Control and Data Acquisition*
- EA* - *Emissão Acústica*
- VIV* - *Vibrações Induzidas por Vórtices*
- NI* - *National Instruments*
- DAQ* - *Data Acquisition*
- LAMEF* - *Laboratório de Metalurgia Física*
- LPS* - *Laboratório de Processamento de Sinais*

1 Introdução

1.1 Contextualização

A expansão da indústria do petróleo tem levado a exploração de grandes profundidades no mar, isso é conhecido como exploração *offshore*. Para transportar esse petróleo extraído, é necessário o uso de tubulações rígidas e flexíveis jacentes no leito marinho, que são submetidas a condições altamente agressivas, sendo afetadas por corrosão, VIV, flambagem radial, tração e compressão. Na Figura 1, mostram-se as estruturas comumente utilizadas na exploração *offshore* no Brasil.

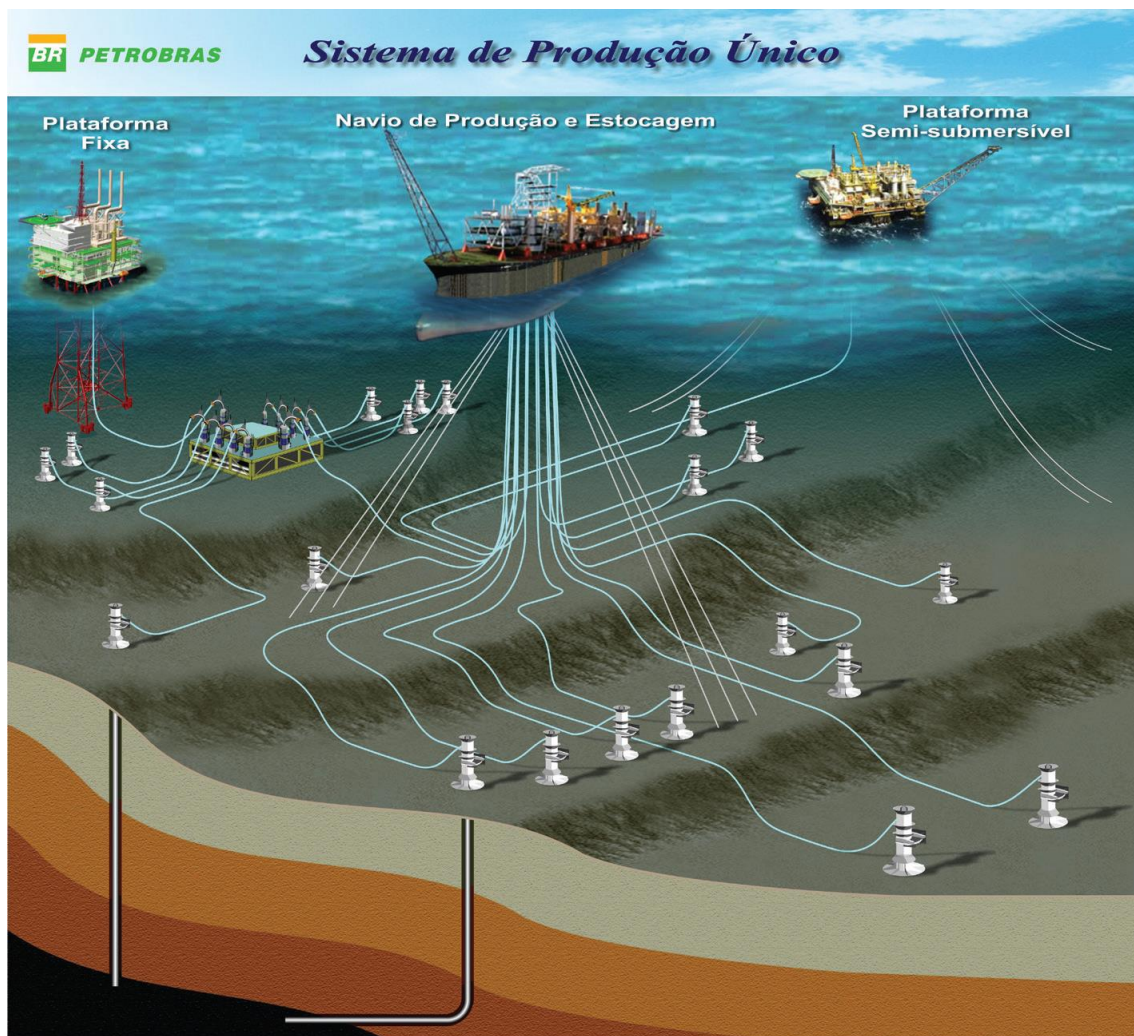


Figura 1. Processo de Produção *Offshore* de Petróleo [8].

Os dutos que conectam o poço com as plataformas são denominados de *risers*, enquanto que os que ficam sobre o solo marinho recebem o nome de dutos submarinos, e que podem ser de dois tipos: *pipelines* e *flowlines*.

Todas as cargas mencionadas anteriormente podem gerar trincas nos dutos rígidos, e que uma vez que aparecem, poderão passar por deformação elástica e plástica e ocasionar a ruptura do duto; produzindo vazamento do petróleo, que é um sério problema para o meio-ambiente.

Nesse cenário, é importante a detecção a tempo dessas trincas, e o reconhecimento da fase na qual se encontram para poder tomar medidas preventivas e evitar um derrame de petróleo. Por isso, que se torna relevante o uso de uma instrumentação adequada para este tipo de ambientes agressivos.

1.2 Problema/Motivação

Desde o início em 2003, a exploração no pré-sal vem crescendo até o momento devido à descoberta de grandes reservas petrolíferas em profundidades de até 3000m de lâmina d'água. Com isto, incrementou-se o uso de dutos submarinos para o escoamento de hidrocarbonetos entre os poços e o continente, e na interligação entre plataformas.

O uso desses dutos rígidos submarinos se deve principalmente à elevada resistência ao colapso em grandes profundidades e a sua simplicidade estrutural. Estes são produzidos sem costura por extrusão, ou com costura de chapas calandradas em forma cilíndrica.



Figura 2. Tubo rígido com costura. Fonte: Palestra de Instalações Submarinos.

Atualmente a manutenção dos *flowlines* e *pipelines* é principalmente do tipo corretiva, onde uma vez que o duto falha pela trinca ou outro tipo de defeito, substituiu-se a seção danificada. Isto ratifica a importância de uma detecção a tempo das trincas e a identificação da fase na qual se encontra, para assim poder tomar medidas preventivas que evitem possíveis vazamentos de petróleo e a contaminação do meio-ambiente; e os altos custos que este tipo de acidentes produzem [15].

Pesquisas anteriores têm se focado na detecção das trincas mediante o uso de sensores de Emissão Acústica, e mediante as técnicas de inteligência computacional foram desenvolvidos classificadores para a identificação das fases das trincas [3].

Porém, ainda não tem se chegado a resultados completamente satisfatórios, nem foi determinado a distância máxima na qual podem ser instalados os sensores EA sem a perda da informação coletada por eles. Tudo isso visando um monitoramento em tempo real dos dutos instalados nas bacias de extração petrolífera.

1.3 Objetivos

Nesse trabalho, pretende-se analisar a forma da onda dos sinais coletados pelos sensores de emissão acústica para classificar a presença de trincas nos dutos rígidos e a fase na qual se encontram, para este caso serão consideradas duas fases: com Propagação Estável (PE) e com Propagação Instável (PI).

Para alcançar o objetivo geral, especificado acima, consideraram-se e se desenvolveram os seguintes objetivos específicos:

- i. Implementar um sistema de aquisição de dados para os sensores de Emissão Acústica utilizando uma placa de aquisição da National Instruments. Isso, com o intuito de obter o sinal original e não os parâmetros já definidos pelos próprios sistemas de aquisição dos sensores EA.
- ii. Desenvolver os classificadores adequados que permitam a identificação das fases na qual se encontram as trincas. Esses classificadores serão obtidos por meio de testes com os dados coletados durante os ensaios no laboratório.

1.4 Justificativa

Os dutos rígidos e flexíveis são altamente usados na exploração do petróleo, tanto na perfuração, quanto na extração e transporte do mesmo para as refinarias. O que os torna elementos cruciais no processo produtivo *offshore*, portanto, também precisam de uma manutenção preventiva e constante para evitar paradas de produção que gerem perdas para as empresas. Porém, a detecção de falhas a tempo destes dutos resulta bastante difícil e normalmente só acontece quando a estrutura já colapsou.

Com o desenvolvimento dos sensores de emissão acústica, abre-se uma oportunidade para a detecção de trincas nos dutos sem que se tenha que parar a produção e de maneira não destrutiva. Mas neste novo cenário será necessário o uso de técnicas de processamentos sinais potentes e de métodos computacionais mais eficazes para que ajudem na tomada de decisões.

2 Revisão Bibliográfica

2.1 Emissão Acústica

Os estudos básicos do fenômeno de emissão acústica têm sua origem a partir de 1948 nos Estados Unidos, mas quem utilizou primeiro a instrumentação eletrônica para registrar o som produzido pela deformação de metais foi Kaiser na Alemanha de 1950 [9].

A partir dos trabalhos de Kaiser, Schofield e Tatro começaram a melhorar a instrumentação nos Estados Unidos; e desde 1964 se dá início a uma série de pesquisas de análise de materiais usando técnicas de emissão acústica (EA) [9].

A EA é um fenômeno que gera onda elásticas transitórias pela rápida liberação de energia de fontes localizadas, que podem vir da geração e propagação de trincas, deformações de material, etc. Para entender melhor, pode-se afirmar que a EA é o campo de tensões criado no material, desta forma, se o campo de tensões não varia não haverá EA.

Quando uma peça de material sólido é submetida a uma carga até a ruptura, produzem-se internamente deslocamentos na microestrutura que podem ser detectados acusticamente. Esse processo vem acompanhado de liberação de energia potencial em forma de calor e de energia vibratória, esta última pode ser detectada na superfície do material, como se pode observar na figura 3.

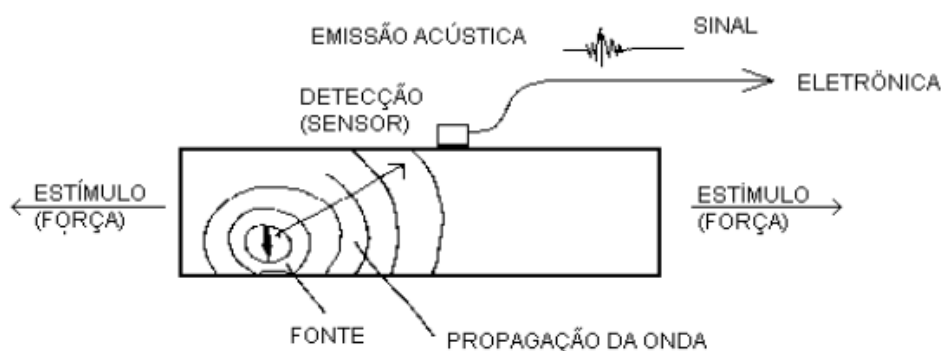


Figura 3. Características da Emissão Acústica [3].

Os sensores devem estar conectados à estrutura, e a saída deles passa por um pré-amplificador e um filtro para tirar qualquer ruído no sinal original. Existem dois tipos

de sensores, com pré-amplificadores externos e integrados; esses últimos foram desenvolvidos para atingir uma alta sensibilidade e para usar cabos longos sem a necessidade de pré-amplificadores separados. Na Figura 4, pode-se observar o diagrama de bloco de um sensor EA integrado.

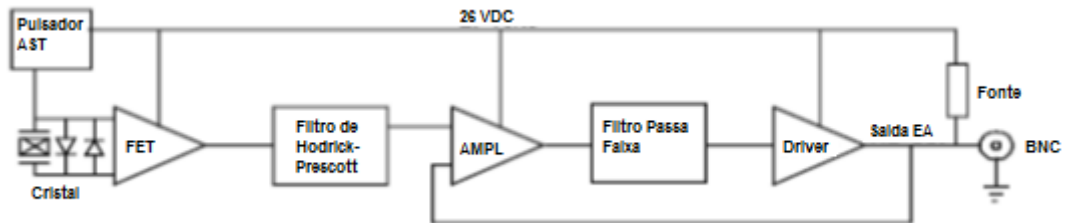


Figura 4. Diagrama de Bloco do Sensor Integrado [6].

2.1.1 Propagação de Ondas Acústicas

Em peças com superfícies planas a onda se propaga em círculos concêntricos ao redor da fonte e pode ser descoberta por um ou vários sensores. Durante a propagação a onda fica atenuada. A distância a qual um evento de EA pode ser descoberto vai depender das propriedades do material, da geometria do corpo de prova, do ambiente, etc. No caso de superfícies planas e cilíndricas, esses eventos podem ser detectados a vários metros de distância.

Nos meios sólidos e elásticos aparecem três tipos de ondas de pressão:

1. Ondas de Compressão (P): Se caracteriza pelo movimento das partículas do meio na direção da propagação. Estas ondas se associam às tensões normais e podem se propagar em qualquer meio. A velocidade de propagação é mais rápida do que outros tipos de onda, e varia em função da geometria, dependendo da relação das dimensões transversais do corpo e do comprimento de onda. No caso em que a dimensão transversal for menor que o comprimento de onda, pode-se aplicar a seguinte fórmula [9]:

$$V_p = \sqrt{\frac{E}{\rho}} \quad (1)$$

Onde:

E é do módulo de elasticidade longitudinal,
 ρ é a densidade.

2. Ondas Cortantes (S): Se caracteriza pelo movimento das partículas do meio na direção perpendicular da propagação. Estas ondas estão associadas às tensões tangenciais e se propagam só em médios sólidos. A velocidade de propagação dessas ondas se expressa mediante a seguinte expressão [9]:

$$V_S = \sqrt{\frac{G}{\rho}} \quad (2)$$

Onde:

G é do módulo de elasticidade transversal,
 ρ é a densidade.

3. Ondas de Superfície (Rayleigh): O movimento das partículas nessas ondas é elíptico retrógrado e as ondas se propagam na superfície do sólido. A velocidade de propagação se estima aproximadamente com a seguinte formulação [9]:

$$V_R = \frac{0,87+1,12\nu}{1+\nu} V_S \quad (3)$$

Onde:

V_S é a velocidade de propagação da onda transversal,
 ν o coeficiente de Poisson.

Por exemplo, para um coeficiente de Poisson de 0,2 a velocidade de Rayleigh seria o 92% da velocidade transversal ou o 56% da velocidade longitudinal [9].

É importante destacar que a sensibilidade dos métodos acústicos para detectar as singularidades do material depende da onda propagada, do comprimento dela e das dimensões das singularidades; sendo que o tamanho do defeito deve ser maior ao comprimento de onda. A relação entre a velocidade da onda, a frequência e o comprimento da mesma é dada por:

$$V = f\lambda \quad (4)$$

Onde:

V é a velocidade da onda,

f é a frequência,

λ é a longitude de onda.

2.1.2 Tipos de Sinais de Emissão Acústica

Os sinais de emissão acústica podem se classificar em:

1. Sinal Transitório (*burst*): Esse tipo de sinais se caracterizam por terem eventos discretos onde os pontos iniciais e finais diferem do ruído de fundo. A presença deste sinal acusa a aparição e crescimento de trincas em um material.

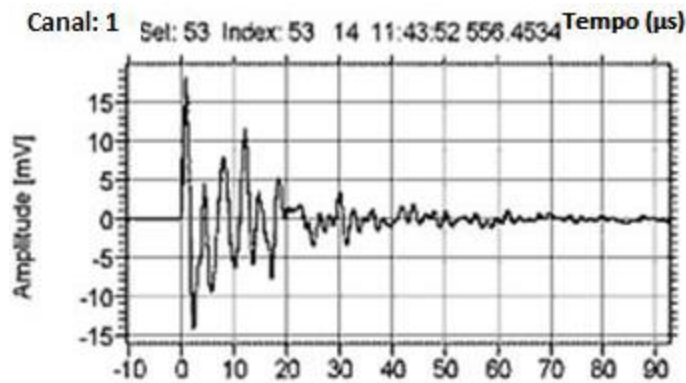


Figura 5. Sinal Transitório [15].

2. Sinal contínuo: São ondas que não têm um ponto final para o sinal, onde só se visualizam diferentes amplitudes e frequências. Estas ondas acusam movimento ou deslocamentos da microestrutura.

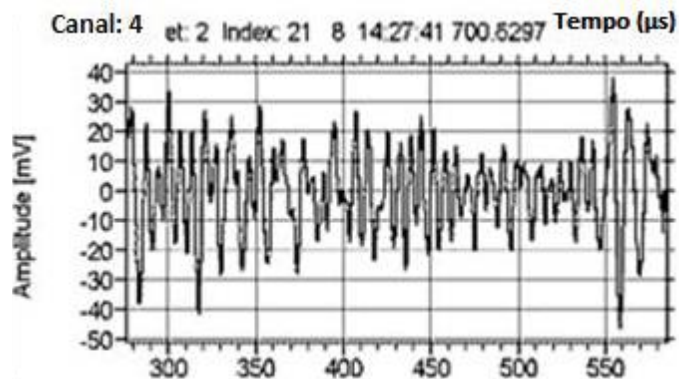


Figura 6. Sinal contínuo [15]

2.1.3 Parâmetros de EA

Da forma da onda é possível extrair informações sobre a fonte, como são a atividade e a intensidade. A seguir se apresenta uma forma de onda típica de EA [9]:

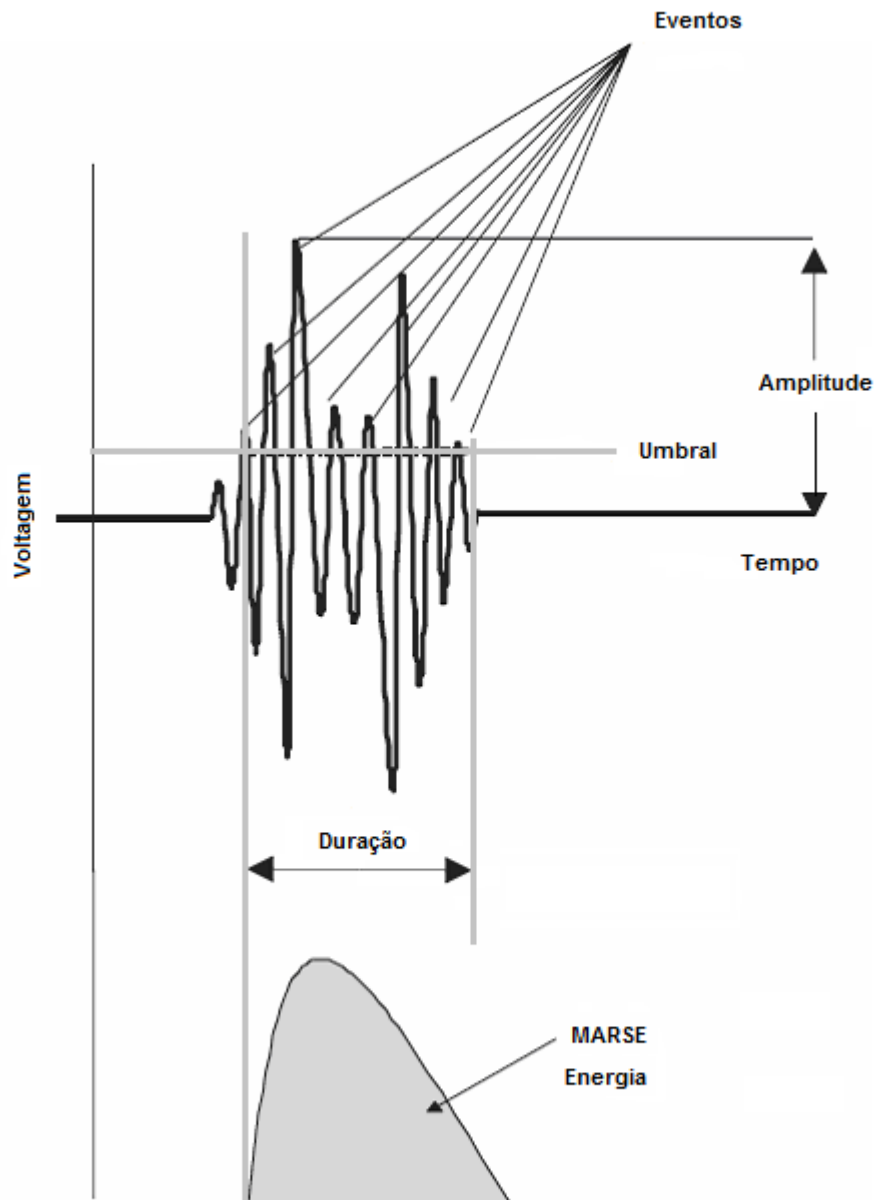


Figura 7. Parâmetros de EA [9]

Na tabela seguinte mostrar-se-ão os parâmetros mais destacados das ondas de EA:

Tabela 1. Principais Parâmetros de EA [9].

Item	Parâmetro	Descrição
1	Tempo de chegada (<i>arrival time</i>)	É o tempo no qual aparece o primeiro sinal que pode ser captado pelos sensores
2	Amplitude Máxima	É o pico máximo que aparece durante a duração do sinal, se expressa em dB.
3	Tempo de subida (<i>rise time</i>)	É o tempo entre a aparição do sinal até atingir a amplitude máxima.
4	Duração	É o intervalo de tempo entre a primeira e última ultrapassagem do limiar do sinal.
5	Contagem	Número de vezes que o sinal ultrapassou o limiar. Vai depender da ressonância do elemento transformador.
6	A-Freq	Frequência média, se obtém com a contagem sobre a duração.
7	MARSE (<i>Measured Area of the Rectified Signal Envelope</i>)	É a área de desenvolvimento do sinal, a energia do sinal de saída do sensor é diretamente proporcional da área contida pela forma de onda e se conhece como força do sinal.

O sinal de EA é afetado por diversos fatores, entre os quais se podem mencionar os da tabela 2 a seguir.

Tabela 2. Fatores que influenciam a EA. [3]

Item	Características da Estrutura e Material	Principais Fatores Favorecendo Sinais de Grande Amplitude	Principais Fatores Favorecendo Sinais de Baixa Amplitude
1	Propriedades Mecânicas	Alta resistência mecânica	Baixa resistência mecânica
2	Estrutura	Anisotropia	Isotropia
		Heterogeneidade	Homogeneidade
		Presença de descontinuidades	Material sem descontinuidades
		Tendência para formar maclas	Muita baixa tendência para maclação
		Transformação martensítica	Transformação de fase por difusão
		Estrutura como fundido	
		Estrutura como soldado	Estrutura relaxada
3	Modo de Ruptura	Crescimento de trinca	Escoamento uniforme
		Ruptura por clivagem	Deformação transversal
4	Modo de carregamento	Alta taxa de deformação	Baixa taxa de deformação
5	Geometria	Grandes espessuras	
6	Ambiente	Baixa temperatura	Alta temperatura

2.1.4 Processamento de Sinais de EA

O processamento de sinais compreende desde a conversão das ondas de EA em um sinal elétrico, a filtragem do ruído, a conversão analógica/digital, até a apresentação dos dados já processados.

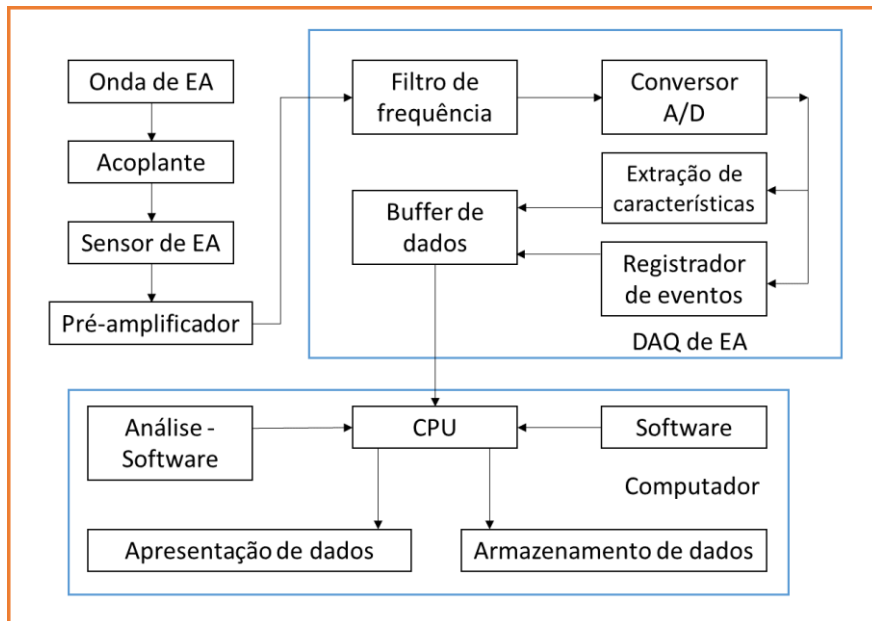


Figura 8. Diagrama de Blocos do Processamento de Sinais [9].

É importante ressaltar que o acoplante é o responsável para assegurar um bom contato acústico entre o sensor e a superfície, normalmente é utilizada uma graxa de silicone ou cola.

Outro fator importante a ter em conta no processamento dos sinais acústicos é que comumente um sinal de EA tem uma duração de $100 \mu s$, o que faz necessário o uso de sistemas DAQ (*Data Acquisition*) com taxas de amostragem na ordem de $1MS/s$ e de uma grande memória e do processador potente.

2.2 Redes Neurais Artificiais

As redes neurais artificiais (RNA) são sistemas de processamento computacional da informação com características semelhantes às redes neurais biológicas, isso se deve a que as primeiras foram inspiradas nas segundas.

O poder de processamento das redes neurais radica na estrutura distribuída paralela, a capacidade de aprendizado e a capacidade de generalizar [10].

As considerações com as que se desenvolveram as redes neurais são as seguintes:

- a. O processamento dos dados é realizado pelos neurônios.
- b. Os sinais se propagam por meio de conexões, que têm associadas a elas um peso para a ponderação do sinal.
- c. A saída é determinada por uma função de ativação associada a cada neurônio.

Portanto, pode-se considerar que uma rede neural se caracteriza pela arquitetura, pelo algoritmo de treinamento e pela função de ativação.

Os métodos de treinamento das redes neurais podem ser classificados em: aprendizado supervisionado e não supervisionado.

2.2.1 Aprendizado Supervisionado

O aprendizado supervisionado se caracteriza por ter nos dados de treinamento padrões de pares de entrada e saída. Devido a que já se conhecem as saídas, o treinamento se vê claramente beneficiado por estar sob a supervisão de um “professor”.

Os parâmetros da rede são atualizados com os vetores de entrada e o erro entre a saída desejada e a saída atual da rede, essa atualização se realiza iterativamente na busca dos parâmetros ótimos.

Para entender melhor esse processo na figura seguinte se mostra um diagrama esquemático deste tipo de aprendizado:

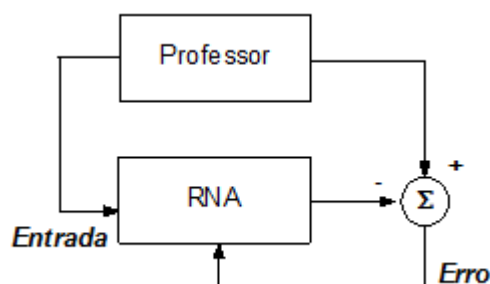


Figura 9. Diagrama de Aprendizado Supervisionado [17].

2.2.2 Aprendizado Não Supervisionado

Para este tipo de treinamento, só são considerados os padrões de entrada, quer dizer que a rede ao não contar com um “professor”, adapta-se baseada nas experiências recolhidas. A seguir se mostra o esquema deste tipo de aprendizado:

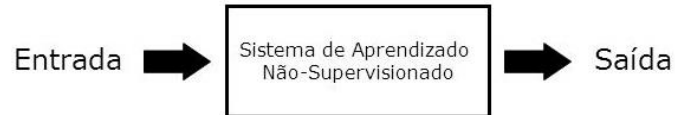


Figura 10. Diagrama de Aprendizado Não Supervisionado [17].

Este tipo de aprendizado é utilizado basicamente pelos seguintes motivos [11]:

- a. Coletar e categorizar um grande conjunto de amostras, que computacionalmente poderiam ter um custo alto.
- b. Treinar um conjunto grande de amostras não categorizadas.
- c. Rastrear mudanças nas características de padrões, que podem ser mudanças lentas.
- d. Encontrar atributos uteis para a categorização.
- e. Conhecer a natureza e a estrutura dos dados em grupos de dados completamente desconhecidos.

Existem vários métodos usados neste tipo de aprendizado, sendo os mais típicos o aprendizado de Hebb, o aprendizado competitivo e o *K-means*, que será explicado a continuação.

2.2.2.1 *K-Means*

O método de *clusterização K-means* é um método não supervisionado utilizado para simplificar o processamento computacional e acelerar a convergência. Consiste em, a partir dos dados não classificados, encontrar as classes (ou *clusters*) com seus respectivos centros de classe de forma iterativa.

Para achar esses grupos, define-se o número de grupos K e se inicializa os centros deles, posteriormente se calculam as distâncias quadráticas euclidianas para deter-

minar a pertencimento de uma entrada a um determinado grupo. Para aprimorar essa classificação inicial, deve-se buscar reduzir essa distância quadrática e se atualizam os centros dos grupos como seus membros.

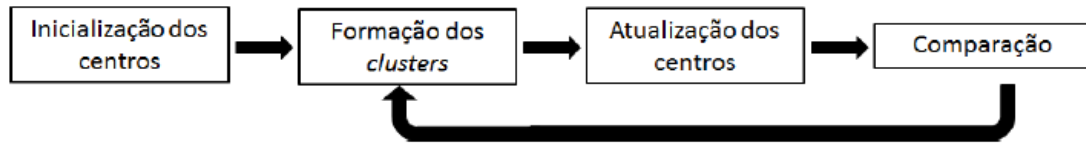


Figura 11. Algoritmo do Método *K-Means* [15].

2.3 Mecânica da fratura em dutos

A mecânica de fratura é um campo que fundamenta seus princípios no comportamento mecânico dos materiais. Podendo classificar os materiais em frágeis ou quase frágeis e em dúcteis[13].

O objetivo da mecânica de fratura é determinar a resistência dos materiais ao avanço de trincas. Na maioria dos casos, a análise está voltada para os materiais dúcteis, porque neles a trinca normalmente tem um crescimento estável quando submetido a cargas externas ou internas antes de colapsar [14].

Existem vários métodos e enfoque para analisar a evolução da trinca nos materiais, tais como a Integral J, CTOD, Curvas R, etc, sendo a Integral J a mais empregada atualmente [14].

A integral J foi derivada inicialmente por Eshelby, mas Rice foi quem começou a aplicá-la na mecânica de fratura. Esta integral de contorno deve envolver a região da trinca, independente do caminho, para ser resolvida [13].

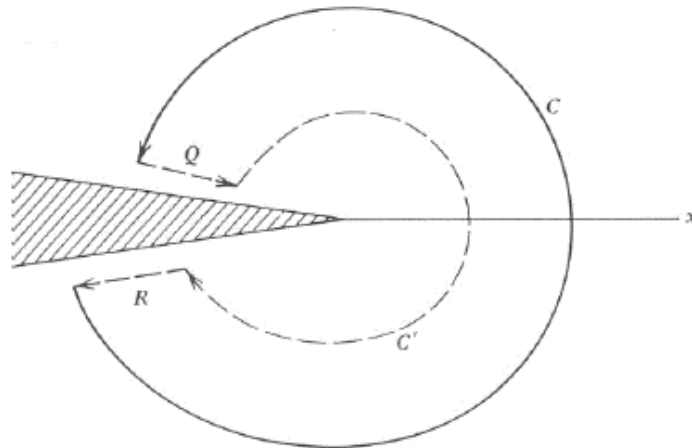


Figura 12. Linha de Contorno em torno da Trinca [13].

A seguir se mostra a curva do crescimento da trinca relacionada à integral J, como se pode observar parece com a curva de deformação de um material.

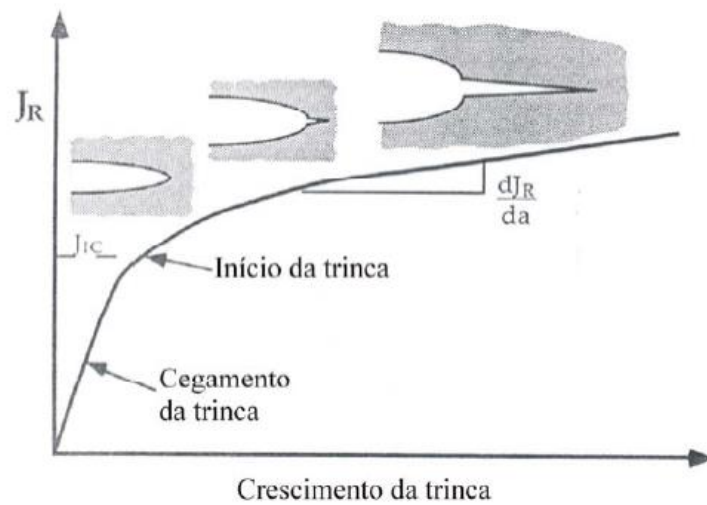


Figura 13. Curva J esquemática para um material dúctil [13].

O conceito físico da integral J é a energia potencial, por unidade de área da trinca, disponível para o crescimento da trinca em um material elástico não linear [14]. Já para o caso linear, a Integral J será equivalente para G.

Consequentemente, pode se considerar que a Integral J é a extensão do conceito de G para o comportamento elástico não linear, dando à seguinte formulação [13]:

$$J = \frac{K_I^2}{E} \quad (5)$$

Onde: K_I é o fator de intensidade crítico de tensões do material,
E é o módulo de Young.

A integral J é considerada como uma candidata a critério de fratura devido a que a integral é independente do caminho escolhido, no caso linear é igual ao parâmetro G e pode ser determinada de forma experimental.

Se o material é submetido a um carregamento que provoque a abertura da trinca, o crescimento da mesma tomara a seguinte forma [13]:

$$J = J_C \quad (6)$$

Para o caso de um estado plano de deformações, o valor crítico estará relacionado com a tenacidade pela seguinte formulação:

$$J_{IC} = \frac{1-v^2}{E} K_{IC} \quad (7)$$

Onde: J_{IC} é o valor crítico no estado plano de deformações de J,

E é o módulo de Young,

v é o módulo de Poisson,

K_{IC} é o fator de intensidade das tensões no estado plano de deformações.

Para o caso de um duto com uma trinca circunferencial passante, como se pode apreciar na figura 14, e submetido a pressão interna, o fator de intensidade de tensões, segundo Takahashi (2002), estará definido pela seguinte formulação [14]:

$$K_I = F_t \sigma_t \sqrt{\pi R_m \theta} \quad (8)$$

Onde: F_t é um fator de forma,

σ_t é a tensão de tração aplicada ao tubo,

R_m é o rádio médio do tubo,

θ é o arco de circunferência em radianes.

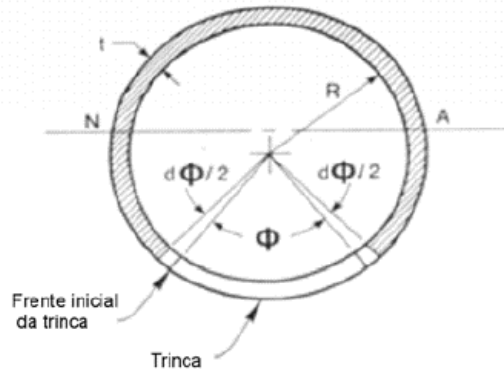


Figura 14. Duto com trinca circunferencial [13].

A tensão de tração aplicada ao tubo está dada por:

$$\sigma_t = \frac{P}{2\pi R_m t} \quad (9)$$

Onde: P é a pressão interna do duto,

t é a espessura da parede do duto.

E o fator de forma tem a seguinte fórmula:

$$F_t = \left[A_t + B_t \left(\frac{\theta}{\pi} \right) + C_t \left(\frac{\theta}{\pi} \right)^2 + D_t \left(\frac{\theta}{\pi} \right)^3 + E_t \left(\frac{\theta}{\pi} \right)^4 \right] \quad (10)$$

Onde: $A_t = 1$,

$$B_t = -1,040 - 3,1831\zeta - 4,83\zeta^2 - 2,369\zeta^3,$$

$$C_t = 16,71 + 23,10\zeta + 50,82\zeta^2 + 18,02\zeta^3,$$

$$D_t = -25,85 - 12,05\zeta - 87,24\zeta^2 - 30,39\zeta^3,$$

$$E_t = 24,70 - 554,18\zeta + 18,09\zeta^2 + 6,745\zeta^3,$$

$$\zeta = \log \left(\frac{t}{R_m} \right).$$

3 Metodologia

Neste capítulo, será explicado como se desenvolveu a pesquisa, a qual constou das seguintes fases experimentais:

1. Implementação do sistema de aquisição de sinais de EA,
2. Desenvolvimento dos classificadores,
3. Análise dos grupos classificados.

A continuação se detalhará cada uma destas fases.

3.1 Implementação do Sistema de Aquisição de Sinais de EA

Nesta fase, foi implementado um sistema de aquisição de sinais para os sensores de emissão acústica, devido aos sistemas de aquisição da PAC só fornecerem os parâmetros da onda, mas não o sinal da mesma. Por isso, tomou-se a decisão de implementar um sistema próprio que permita obter a forma de onda do sinal EA para uma análise diretamente da fonte do ensaio.

3.1.1 Equipamentos

Para a implementação do sistema DAQ foram usados os equipamentos da *National Instruments (NI)*, os sensores foram os da PAC e para o processamento computacional foi utilizada a plataforma *LabView*, que também é da *NI*.

Para dimensionar os equipamentos, devem-se ter em conta o processo que se deseja mensurar, e dado que no LAMEF estavam utilizando os sensores R15I-AST, decidiu-se continuar trabalhando com eles e tomá-los como referência para o dimensionamento. Mas também se deve entender como que o ensaio era realizado e como que acontece o fenômeno de emissão acústica.

Como já se mencionou anteriormente, o período de um sinal de EA está na ordem dos microssegundos, em consequência, a frequência está na ordem dos mega-

Hertz. E de acordo a especificação técnica do sensor a frequência de operação dele está no *range* de 80 a 200KHz.

Para evitar o vazamento espectral no processamento de sinal, a frequência mínima de aquisição deve ser de 2,5 vezes a frequência de operação (a frequência do sinal real). Portanto, considerando o valor máximo do range do sensor, a frequência de aquisição deveria ser no mínimo de 500KHz.

Com essas informações, selecionou-se a placa de aquisição adequada para o ensaio, sendo está a placa NI PCI-6250 da NI, cuja frequência de aquisição é de 1MS/s no modo multicanal (essa frequência é multiplexada entre todas as entradas) ou de 1,25MS/s usando um canal só. Além disso, a placa conta com 16 entradas analógicas.

Dado que a placa selecionada era do tipo PCI, também foi necessário usar um módulo conector de terminais para conectar a saída dos sensores. A seguir se mostram as fotos dos equipamentos utilizados.

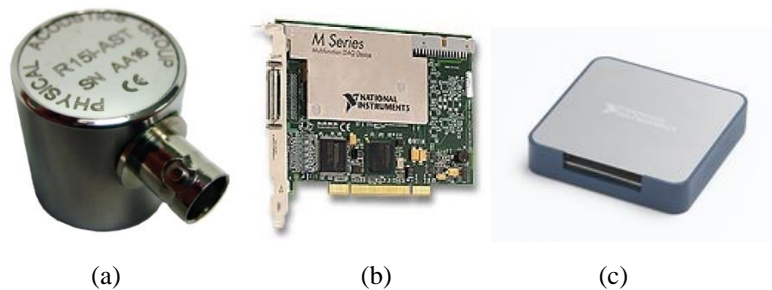


Figura 15. Equipamentos de Aquisição: (a) Sensor R15I-AST, (b) Placa NI PCI-6250 e (c) Bloco conector SCB-68A.

Dado que o bloco conector é do tipo borne e os conectores dos sensores são coaxiais, foi necessário fazer uma adaptação do conector coaxial para borne e além incluir o circuito de alimentação conforme especificação do sensor de EA.

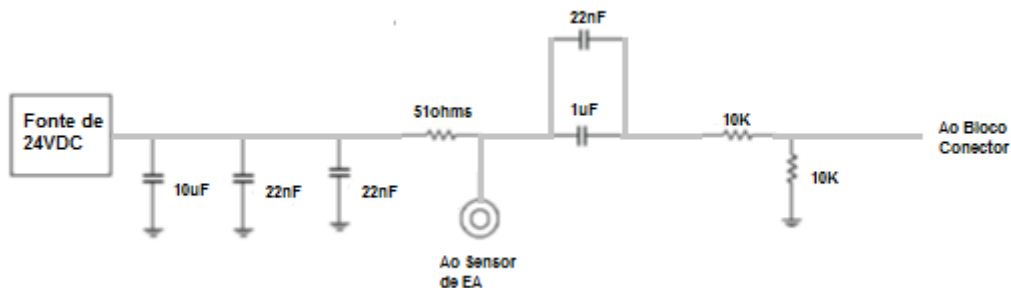


Figura 16. Circuito de conexão dos sensores com o sistema de aquisição.

3.1.2 Software de Aquisição

Com os equipamentos prontos para a aquisição de sinais, continuou-se com o desenvolvimento do programa na plataforma *LabView* 2014. Além de ser instalado o programa, é necessário instalar os drivers da placa e o *NI-DAQmx*. Neste último, realizam-se as primeiras configurações do *hardware* e se verifica se o mesmo está sem defeitos. É importante salientar que sem realizar essas configurações o *LabView* não reconhecerá as entradas dos sensores.

Basicamente o programa foi desenvolvido em duas partes:

- 1) Leitura dos sinais nas entradas e visualização gráfica online;
- 2) Gravação dos sinais adquiridos.

A NI desenvolveu um formato para armazenar dados próprios da instrumentação de forma que sejam mais compactos do que os outros formatos. Esse novo formato é o TDMS, cujas siglas em inglês são *Technical Data Management Streaming*. Para poder abrir essa data e processá-la posteriormente pode-se usar o *Diadem*, também da NI.

Na figura 16, mostra-se a interface do programa para adquirir os sinais e na figura 17 se mostram os dados que foram gravados sendo visualizados no *Diadem*.

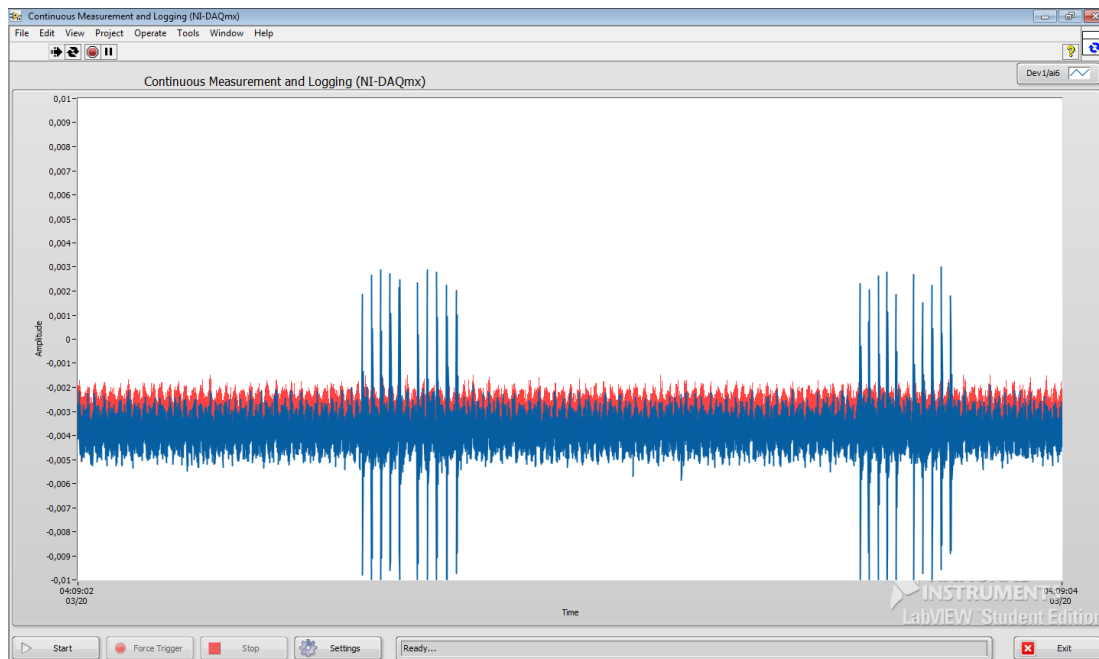


Figura 17. Interface do Sistema de Aquisição de Sinais de EA.

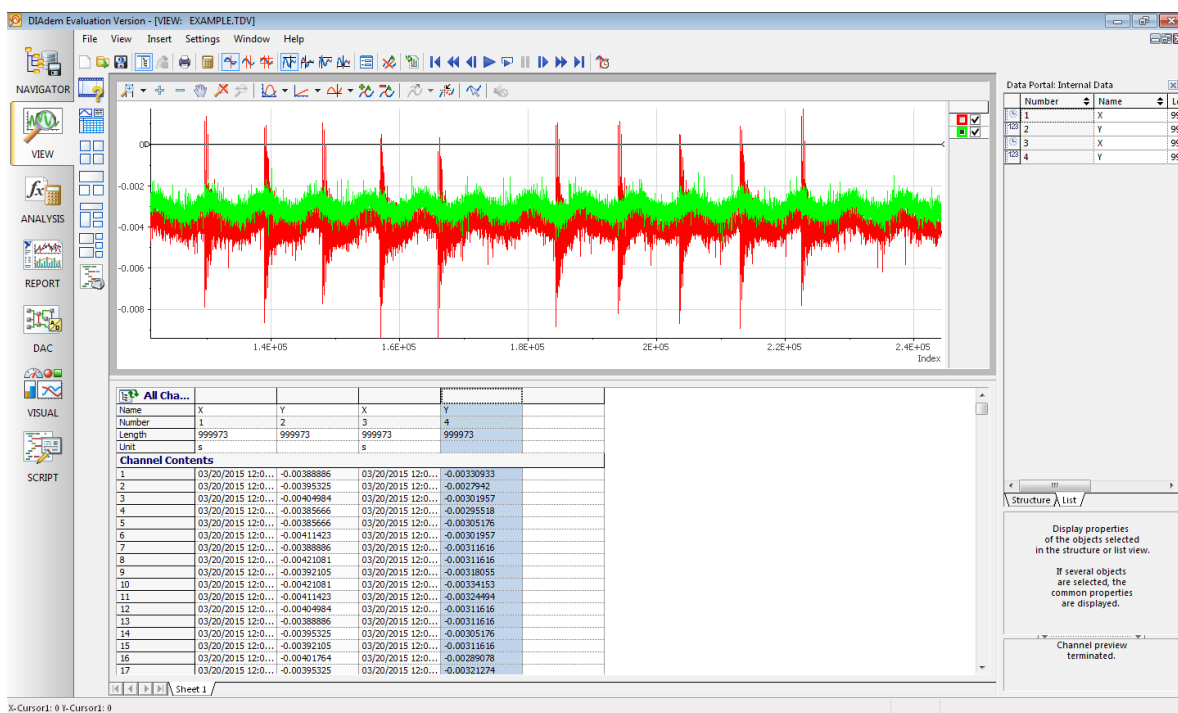


Figura 18. Interface Diadem.

3.1.3 Descrição do Ensaio

Esta pesquisa faz parte do projeto “ Inspeção de Dutos Rígidos” entre o CEN-PES (Petrobras), o Laboratório de Metalurgia Física (LAMEF) da UFRGS, e o Laboratório de Processamento de Sinais (LPS) da UFRJ. O projeto consiste na realização de ensaios, nos quais um duto de 40 m de comprimento com trincas usinadas de forma elípticas na superfície é pressurizado.



Figura 19. Duto com sensores de EA montados.

A pressurização do duto se realiza em várias fases desde o início do carregamento com água até o colapso da estrutura, alternando fases de aumento de pressão com patamares de pressão constante.

O crescimento em profundidade da trinca inserida se mede mediante a técnica ToFD, cujas siglas em inglês são *Time of Flight Diffraction* (Tempo de Percurso da Onda Difrata). Esta técnica consiste na estimação do tempo de voo das ondas de ultrassom difratadas nas pontas de discontinuidades, que é diretamente proporcional à posição real delas. Para isso, utilizam-se um emissor e um receptor [15].

Quando se emite a onda de ultrassom, está poderá ser difratada devido à presença de algum defeito no material no qual se propaga, neste caso, as ondas difratadas da técnica ToFD são utilizadas para medir a profundidade da trinca inserida artificialmente

na superfície da parede interna do duto. Com o sinal coletado, é possível reconstruir uma imagem bidimensional do defeito [15].

Além dos transdutores ultrassônicos, foram utilizados 10 sensores de EA, cujo modelo é o R15I-AST da PAC, e para aquisição de sinais o Disp 16C da PAC também. Os parâmetros que o equipamento adquire e disponibiliza são apresentados na Tabela 3.

Tabela 3. Parâmetros de EA adquiridos pelo sistema PAC [3].

Item	Parâmetro	Descrição
1	<i>Rise</i>	Tempo desde que o primeiro pico ultrapassa o limiar até atingir a amplitude máxima.
2	<i>Count</i>	Número de vezes que o sinal ultrapassou o limiar.
3	<i>Energy</i>	Integral do módulo de cada pico de onda. A amplitude se expressa em volts e não em decibel.
4	<i>Duration</i>	É o intervalo de tempo entre a primeira e última ultrapassagem do limiar do sinal.
5	<i>AMP</i>	Amplitude máxima do sinal, medido em dB.
6	<i>A-Freq</i>	Frequência média, se obtém com a contagem sobre a duração.
7	<i>RMS</i>	Raiz da integral dos valores de tensão ao quadrado dividido pelo intervalo de tempo dado pela diferença de tempo entre os limites superior e inferior da integração.
8	<i>ASL</i>	Nível médio do sinal, medido em Db.
9	<i>PCNTS</i>	Número de picos até a amplitude máxima ser alcançada.
10	<i>R-Freq</i>	Frequência de reverberação: [contagem – contagem ao pico] / [duração - tempo de subida].
11	<i>I-Freq</i>	Frequência de iniciação: contagem ao pico / tempo de subida.
12	<i>Sig-Strength</i>	Potência do sinal.
13	<i>ABS-Energy</i>	Energia absoluta, definida como a quantidade real da energia.
14	Freq-PP1	Primeira potência parcial.
15	Freq-PP2	Segunda potência parcial.
16	Freq-PP3	Terceira potência parcial.
17	Freq-PP4	Quarta potência parcial.
18	C-Freq	Centroide de frequência.
19	P-Freq	Frequência pico, que a de maior magnitude no espectro.

Como complemento ao ensaio se utilizará também o sistema de aquisição implementado para esta pesquisa, com o intuito de obter a forma de onda gerada pelo sensor e depois processá-la.

Devido às informações pouco detalhadas no *datasheet* da placa NI PCI-6250, no momento da compra pensou-se que a taxa de amostragem de 1MS/s era para cada canal da placa. Porém, durante a implementação observou-se que essa taxa era multiplexada entre todos os canais da placa; quer dizer, que só foi possível monitorar dos sensores de EA com uma taxa de aquisição de 500KHz.

3.1.4 Definição das Classes de trinca

Com a pressão do duto monitorada e com o sinal de ultrassom coletado se pode estabelecer as fases pelas quais passam as trincas. Da teoria de resistência de materiais e da mecânica de fratura, sabe-se que as deformações de um corpo de prova ou de um defeito passam pelas seguintes fases:

1. Fase elástica: Nessa fase o material deformado retorna a sua forma original quando o carregamento é interrompido.
2. Fase plástica: Nessa fase o material deformado não retorna a sua forma original depois de que a carga cessa.

Seguindo esse mesmo raciocínio, agruparam-se os dados coletados nas seguintes classes:

1. Propagação Estável (PE): Nessa classe, consideram-se os sinais que não apresentam propagação da trinca e também aqueles que têm um crescimento lento da trinca.
2. Propagação Instável (PI): Nessa classe, encontram-se os sinais cuja propagação da trinca se dá de forma instável até o momento que o duto colapsa.

3.1.5 Pré – Processamento dos dados

Dado que o sistema de monitoramento da pressão no duto rígido é independente do sistema de EA, é necessário de sincronizar os dados coletados no tempo.

Nessa pesquisa, trabalhar-se-á diretamente com a forma de onda dos sinais de emissão acústica, por isso, como parte do pré-processamento se aplicou a transformada de Fourier (FFT) a cada sinal coletado. E depois se normalizaram as transformadas para aumentar o desempenho do classificador durante o treinamento.

3.2 Desenvolvimento do Classificador

Devido a alguns problemas técnicos durante o ensaio realizado no LAMEF, usando o sistema de aquisição com equipamentos da *NI* e que serão explicados no próximo capítulo, não será possível utilizar os dados coletados durante o mesmo.

Destarte, para o desenvolvimento dos classificadores nesta fase foram utilizados os dados do Ensaio 3 dos 6 realizados para o projeto “Inspeção de Dutos Rígidos” da UFRGS [3].

3.2.1 Breve descrição do Ensaio 3

Os dados foram coletados com o sistema descrito anteriormente no ponto 3.1.3. E para adquirir a forma de onda foram instalados, paralelamente aos sensores da PAC, 6 sensores de EA da Vallen. A distribuição de todos os sensores sobre o duto de prova se mostra na seguinte figura.

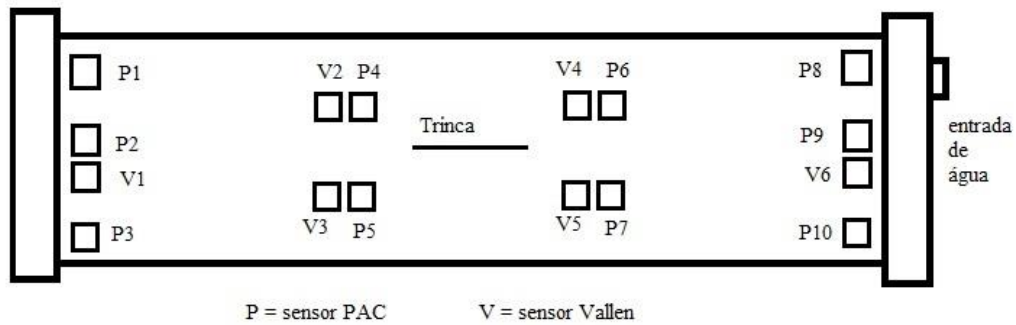


Figura 20. Distribuição dos sensores no duto.

Observa-se que a trinca se encontra na metade do duto, e a distribuição dos sensores segue o seguinte formato:

Tabela 4. Quantidade dos sensores EA sobre o duto.

Item	Posição	Sensores PAC	Sensores Vallen
1	Extrema Esquerda	3	1
2	Centro Esquerda	2	2
3	Centro Direita	2	2
4	Extrema Direita	3	1

O ensaio teve uma duração de aproximadamente 4 horas, desde o início da pressurização até o colapso do duto. E neste caso a taxa de amostragem foi de 1MHz, desta forma se coletou uma grande quantidade de dados.

Foi necessário fazer uma de seleção dos dados que seriam utilizados para nossa pesquisa, e o primeiro filtro foi extrair os sinais monitorados durante a fase de enchimento do duto com água. O segundo filtro foi desconsiderar a fase inicial de pressurização e a estabilização da pressão.

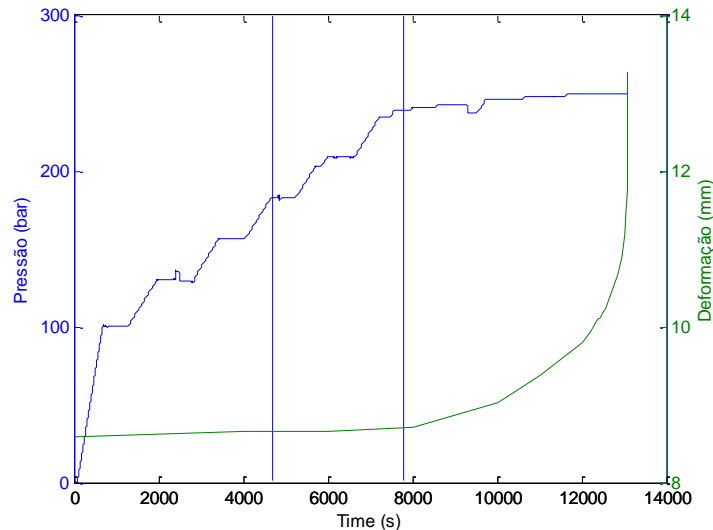


Figura 21. Curva de Pressão e Deformação vs tempo.

Como se observa na figura anterior a trinca se encontra na fase Sem Propagação (SP) desde o início até os primeiros 6500 segundos, a partir de aí começa a fase de Propagação Estável (PE) até os 8000 segundos; onde começará a Propagação Instável (PI) da trinca até o colapso do duto, que foi aos 13000 segundos.

Para este trabalho consideraram-se os dados coletados entre a última fase PE, a transição de PE a PI e a parte inicial da fase PI. Portanto, trabalhar-se-á com dois conjuntos de dados, o primeiro que inclui 36 formas de onda; onde as fases PE e PI são diferenciadas. E no segundo conjunto com 68 formas de onda onde se inclui a transição entre as fases PE e PI.

3.2.2 O Classificador

Para o desenvolvimento do classificador, foram escolhidos dois conjuntos de dados, um menor com um total de 18 formas de onda referentes à fase PE e 18 à fase PI; e um segundo conjunto composto de 34 formas de onda referentes à fase PE e 34 à fase PI.

O método selecionado para o agrupamento dos dados foi o método de *clusterização K-means*, devido à facilidade de implementação no *MatLab* e ao baixo custo computacional.

Para cada conjunto, consideraram-se dois tipos de inicialização dos centroides, sendo estes: inicialização aleatória e definindo os centroides inicialmente baseados na média das formas de onda de PE e de PI.

Em cada um dos tipos de inicialização, tomou-se como premissa a divisão do conjunto de dados em duas classes. Após essa primeira classificação, procedeu-se a subdividir aquelas classes onde ainda existiam tantos dados de PE como de PI, seguindo a estrutura mostrada na figura 22.

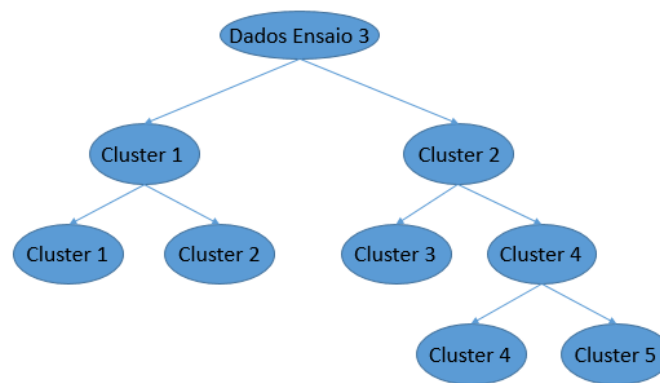


Figura 22. Classificação com estrutura tipo árvore.

No processo de subdivisão de classes, foram consideradas duas formas, a primeira onde se manteve o centroide e os sinais de somente um tipo (seja PE ou PI) ou no caso em que tivesse um ou dois elementos diferentes a da maioria da classe. E depois subdividir em outras duas classes, aqueles grupos onde se tinha mais de duas formas de ondas PE e PI. Esta primeira forma encontra-se retratada também na Figura 22.

Na segunda forma de subdivisão, só se mantiveram os centroides daquelas classes que já não podiam ser mais subdividas, e realizando a *clusterização* do conjunto total de dados, tal e como apresenta-se na figura seguinte.

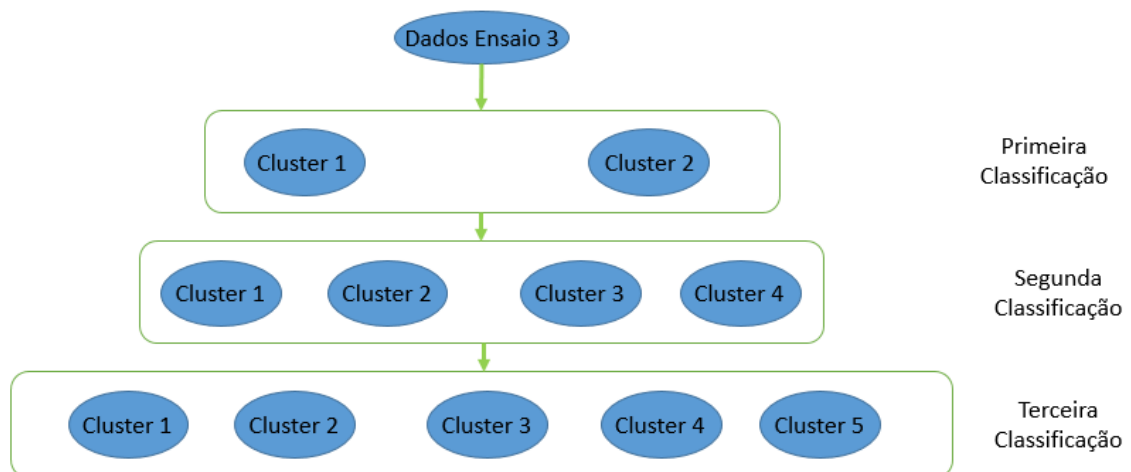


Figura 23. Classificação *K-Means* reclassificando o total de dados.

Desta forma, com as considerações mencionadas acima ter-se-á um total de quatro casos para cada conjunto de dados, e cujos resultados serão apresentados no próximo capítulo.

3.3 Análise dos grupos classificados

Após realizar a classificação dos dados, tanto para o conjunto de 36 como o de 68 formas de onda, é importante realizar uma análise do agrupamento para verificar o grau de acurácia do classificador.

Neste trabalho, a análise de acurácia do classificador constou das seguintes fases:

1. Fase 1: Foi analisada o índice de acerto do classificador final, considerando o conjunto total de dados.
2. Validação Cruzada: Para o segundo conjunto de dados, aplicou-se o método de validação cruzada. Onde foram considerados 3 subconjuntos, com 22 formas de onda cada um, dos quais dois subconjuntos foram utilizados para o treinamento e um para o teste/validação. Foram estimados os índices de acerto para cada uma

das combinações de subconjuntos, e depois calculou-se o índice médio de acerto.

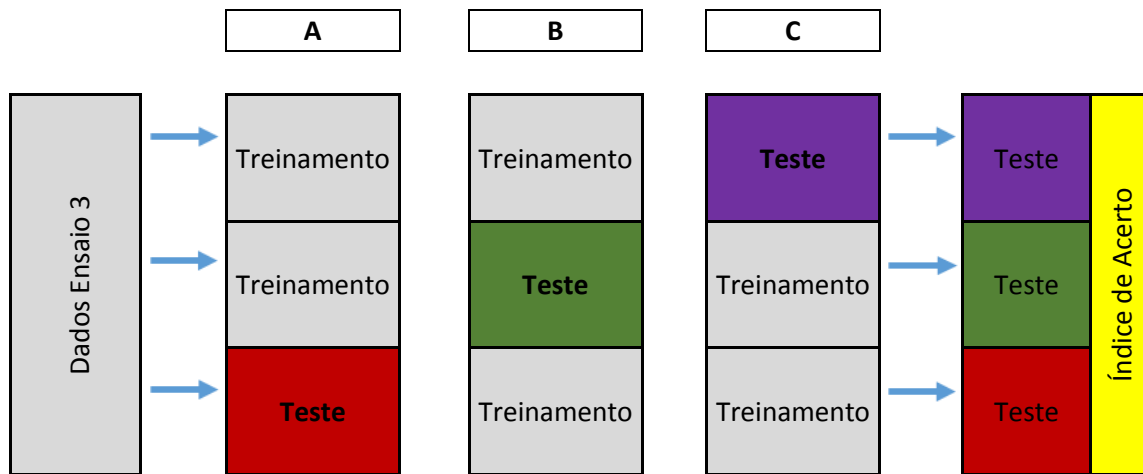


Figura 24. Validação Cruzada.

3. Fase 3: O conjunto de 68 dados foi utilizado como um conjunto de teste para avaliar a acurácia dos grupos obtidos durante a classificação do conjunto de 36 dados. Para isso, foram utilizados os centroides obtidos para cada caso considerado.

4 Resultados

Neste capítulo, serão apresentados os resultados e suas respectivas análises.

Este capítulo está estruturado da seguinte forma:

1. Na seção 4.1, explicar-se-ão alguns detalhes da implementação do sistema de aquisição, e as dificuldades achadas durante o processo.
2. Na seção 4.2, serão apresentados os resultados obtidos para a classificação do conjunto de 36 dados. E em cada subseção desta mostrar-se-ão os resultados dos casos considerados.
3. Na seção 4.3, terá a mesma estrutura que a seção 4.2, mas utilizando o conjunto de 68 formas de onda.
4. Na seção 4.4, utilizar-se-á o método validação cruzada sobre o conjunto de dados com 68 formas de onda.
5. Na seção 4.5, será verificada a acurácia dos classificadores desenvolvidos na seção 4.1.

4.1 Resultados da Implementação do Sistema de Aquisição de Sinais de EA

Conforme se detalhou no capítulo anterior, foi implementado um sistema de aquisição de sinais usando uma placa PCI com uma taxa de amostragem de 1MS/s multiplexada da NI e a interface gráfica foi desenvolvida na plataforma LabView.

A versão do software era uma para estudante e com uma duração dos privilégios completos de 6 meses, após os quais os *toolkits* utilizados seriam limitados de forma compatível com a licença.

Durante a implementação do sistema, foi identificado que a alimentação de 24VDC dos sensores utilizados da PAC usava o mesmo cabo coaxial pelo qual era coletado o sinal de EA. Conseqüentemente, foi necessário realizar algumas adaptações para alimentar o sensor e derivar o sinal de saída para a placa PCI.

Estas adaptações basearam-se no seguinte circuito, especificado pelo fabricante no *datasheet* do sensor:

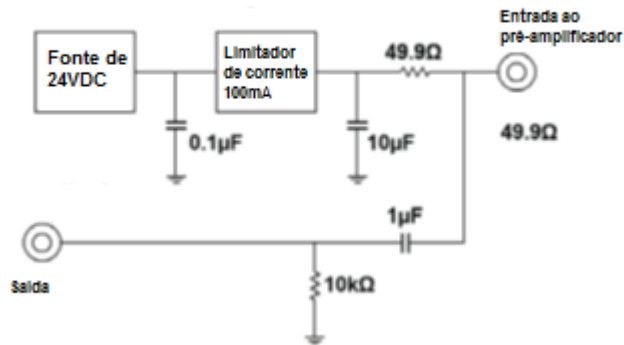


Figura 25. Circuito de Alimentação do Sensor [5].

No *datasheet* do sensor, especifica-se que a alimentação do sensor deve ser entre 20 e 30VDC, por isso foi utilizada uma fonte de 24VDC. Porém, a placa PCI da NI só trabalha com sinais de até máximo 10VDC.

A medida tomada para poder utilizar o sensor com a placa foi a de implementar um circuito de divisão de voltagem para que a saída obtida do sensor esteja dentro dos limites tolerados pela placa.

O desenvolvimento da interface no *LabView* foi mais simples, devido a facilidade de manipulação dos *toolkits* gráficos da plataforma. Cabe salientar que é importante a configuração da placa no MAX, para facilitar a leitura dos canais na interface gráfica.

Como já se mencionou anteriormente, os dados adquiridos foram salvados no formato *tdms*, pelo nível de compactação e a facilidade de manipulação depois em diferentes programas, como por exemplo no *MatLab* ou no *Microsoft Excel*.

Durante o ensaio realizado no LAMEF em março de 2015 foi utilizado o sistema implementado. Porém, os resultados não foram satisfatórios por dois motivos. O primeiro motivo foi que o duto de 40m não colapsou, porque a trinca usinada nele não tinha

profundidade suficiente para leva-la à propagação instável e consequente fratura do duto. O segundo motivo foi que o ensaio teve uma duração de aproximadamente 6 horas, gerando-se uma grande quantidade de dados. O arquivo gerado ficou com tamanho de 300GB e mais de dois bilhões de medições, por isso, foi impossível abrir o arquivo no *Diadem*, *Excel* ou no *MatLab*.

Esse tamanho de dados é considerado de megadados, e para poder processá-la é necessário o uso de *clusters*. Ante essa dificuldade, tentou-se usar um programa, desenvolvido por um membro do foro da NI, para particionar os megadados armazenados em formato *tdms*. Porém, não foi possível realizar esta partição em arquivos menores, devido à limitação da memória *RAM* do computador para poder alocar o arquivo original. Diante do exposto, foi decidido usar as formas de onda adquiridas no Ensaio 3 realizado anteriormente pelo LAMEF com um duto de 2m, cuja instrumentação mostra-se na Figura 26.



Figura 26. Duto de 2m de comprimento instrumentado [18].

4.2 Classificação usando um conjunto de 36 dados

Nesta seção, a classificação será realizada utilizando o método não supervisionado *K-means* para o conjunto de 36 formas de onda. Estes sinais foram escolhidos baseados na curva de pressão e da deformação da trinca ao longo do ensaio, conforme foi visto na Figura 21 do capítulo 3.

Este conjunto de dados é formado por 18 sinais de PE e 18 sinais de PI, onde os sinais PE foram adquiridos entre os 4700 e 6800 segundos de duração do ensaio. E os sinais PI foram adquiridos a partir dos 7800 segundos do ensaio até o colapso da amostra.

O total de sinais de PI presente no ensaio 3 é maior do que os sinais PE, aproximadamente 150 sinais. Por esse motivo, só se escolheram 18 desse total para ter uma quantidade igual de sinais de PE e de PI. As formas de onda de PI foram escolhidas uma a cada 8 sinais coletadas no tempo.

Na figura 27, apresentam-se a forma onda de um sinal de EA da trinca com propagação estável e a sua respectiva transformada de Fourier, e na Figura 18 a da trinca com propagação instável.

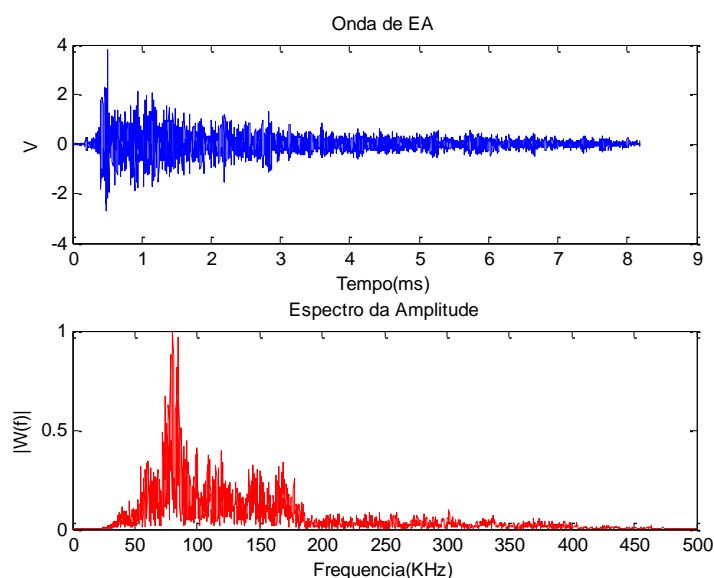


Figura 27. Forma de Onda PE.

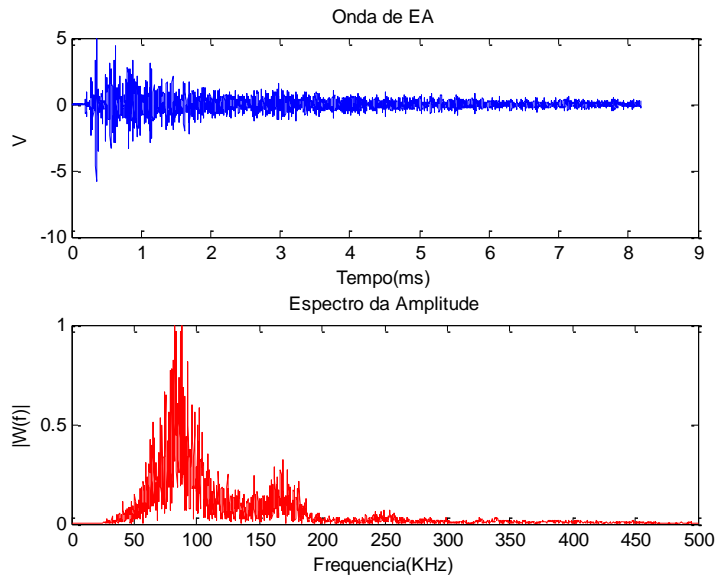


Figura 28. Forma de Onda PI.

Como já foi explicado na seção 3.2.2, o *K-means* será aplicado primeiro ao conjunto total de dados, considerando duas classes e depois serão subdividas as classes que tenham a presença de mais de duas de formas de onda de propagação estável e mais de duas de propagação instável. Esse tipo de subdivisão será denominado como classificação tipo árvore.

Além da forma de subclassificação, tiveram-se em consideração a inicialização aleatória dos centroides das classes e a inicialização dos centroides, estimando a média das formas de onda de PE e a média das ondas de PI.

Também foram considerados mais dois casos, que denominar-se-ão sem reclassificação e com reclassificação. O primeiro atende ao caso explicado anteriormente de manter a classe cujos elementos só pertençam à PE ou à PI e reclassificando as classes que ainda podem ser divididas em duas subclasses.

O caso com reclassificação é aquele no qual manter-se-á o centroide da classe que já não pode ser dividida e reclassificando o conjunto total de dados.

É importante explicar que para todos os casos nos quais se subdividirá alguma classe, estimar-se-ão os centroides calculando as medidas dos elementos que são PE e dos que são PI.

Foi considerado como critério de parada da subdivisão de classes um índice de acerto maior a 80%, como nos resultados obtidos na pesquisa do Carlim [3]. Além disso, foi considerado um máximo de um ou dois intrusos em qualquer uma das classes definidas.

4.2.1 Caso 1: Classificação Árvore sem reclassificação

Para a classificação usando o método *K-means*, utilizou-se como entrada a transformada de Fourier das formas de onda dos sinais de EA, e dado que todas as formas de onda geradas estão compostas por 8192 amostras no tempo também a FFT terá um total de 8192 no espectro da frequência.

Como a transformada tem o efeito de espelhar a amplitude da onda na frequência, só se trabalhou com a metade das amostras, quer dizer que cada entrada do classificador terá um total de 4097 parâmetros. Consequentemente, a matriz de entrada terá uma dimensão de 36x4097.

Para poder plotar o conjunto de entradas, e posteriormente elas já classificadas, em um plano 2D, tomou-se como referência a frequência onde a amplitude é máxima e a média de todas as amplitudes no espectro. Na continuação apresenta-se os dados PE e PI plotados seguindo a regra anterior:

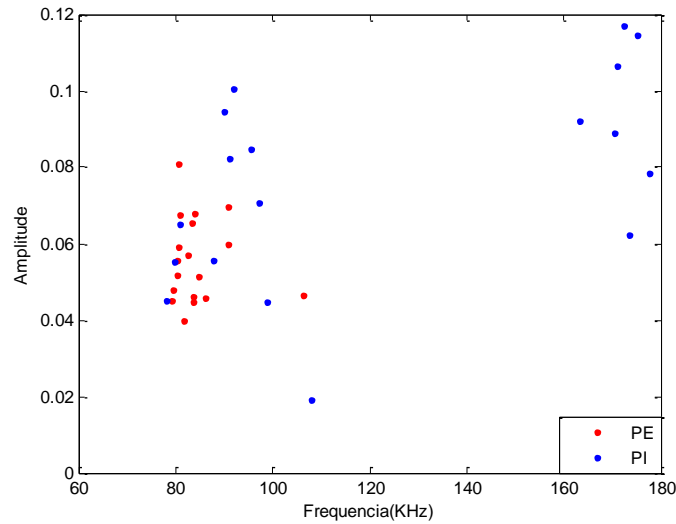


Figura 29. Entradas PE e PI.

Neste primeiro caso, com a matriz de entrada definida se aplicou a classificação *K-means*, considerando a distância quadrática euclidiana, centroides aleatórios e inicialmente uma partição em duas classes. O resultado obtido é apresentado na figura seguinte:

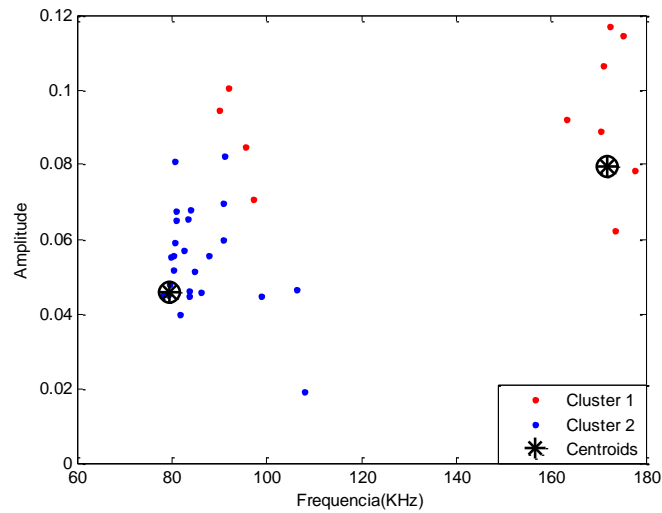


Figura 30. Classificação com 2 classes para o Caso 1.

Na Figura 30, observa-se que o *Cluster 1* tem 11 elementos e o *Cluster 2* tem 25, para ver o índice de acerto desta primeira classificação apresenta-se a seguinte tabela:

Tabela 5. Índice de Acerto para o Caso 1 com duas classes.

Item	Classe	Ondas PE	Ondas PI
1	Cluster 1	0	11
2	Cluster 2	18	7
Índice de Acerto		80,56%	

Portanto, manter-se-á o *Cluster 1* com seus 11 elementos e centroide. O *Cluster 2* será dividido em duas novas classes e os centroides dessas classes serão a média das 18 ondas de PE e a média das 7 ondas de PI. Os resultados obtidos mostram-se na Figura 31 e na Tabela 6.

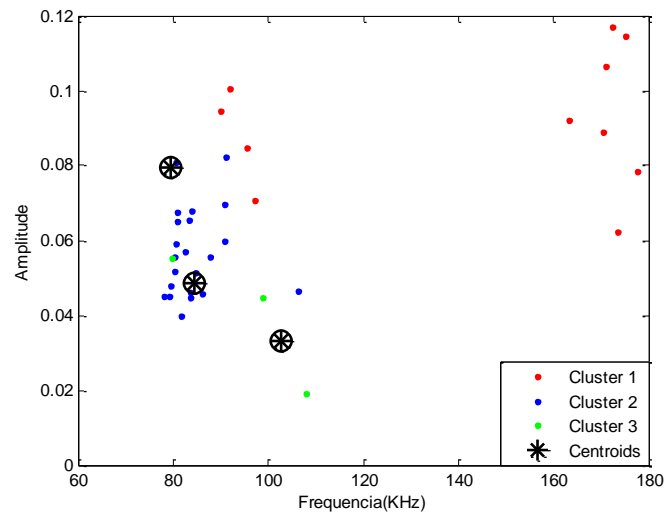


Figura 31. Classificação com 3 Classes para o Caso 1.

Tabela 6. Índice de Acerto para o Caso 1 com 3 classes

Item	Classe	Ondas PE	Ondas PI
1	Cluster 1	0	11
2	Cluster 2	18	4
3	Cluster 3	0	3
Índice de Acerto		88,89%	

Da Tabela 6, observa-se que tanto o *Cluster 1* como o *Cluster 3* somente têm elementos de PI, assim, essas duas classes se manterão. No caso do *Cluster 2*, ainda existem elementos de PI, por esse motivo se dividirá em duas novas classes. Os resultados desta nova divisão do *Cluster 2* mostram-se na Figura 32 e na Tabela 7.

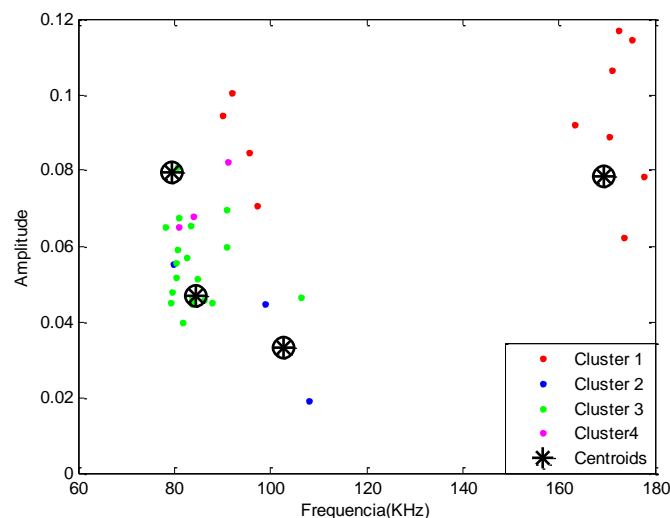


Figura 32. Classificação com 4 Classes para o Caso 1.

Tabela 7. Índice de Acerto para o Caso 1 com 4 Classes.

Ítem	Classe	Ondas PE	Ondas PI
1	Cluster 1	0	11
2	Cluster 2	1	2
3	Cluster 3	17	2
4	Cluster 4	0	3
Índice de Acerto		91,67%	

Da Tabela 7, nota-se que o *Cluster 1* e o *Cluster 4* são os que se mantiveram da classificação anterior. Os *Clusters 2* e *3* ainda têm dois ondas PI, porém, já não serão subdivididos, porque o índice de acerto é alto e ao continuar subdividindo estar-se-á reduzindo a capacidade de generalização do classificador. E também, ao aumentar o número de classes pode-se ocasionar um super-treinamento do classificador.

Ao encerrar a classificação neste ponto será considerado que o *Cluster 1*, o *Cluster 2* e o *Cluster 4* são os que contêm as ondas de PI, enquanto que o *Cluster 3* contém as ondas PE.

4.2.2 Caso 2: Classificação árvore sem reclassificação e inicializando os centroides

Neste segundo caso, tomaram-se as mesmas considerações que no Caso 1, com a única diferença de que ao invés de inicializar os centroides aleatoriamente para a classificação em duas classes, foram utilizados dois centroides calculados pela média das ondas de PE e das ondas de PI.

Os resultados obtidos foram plotados na Figura 33 e o índice de acerto é detalhado na Tabela 8.

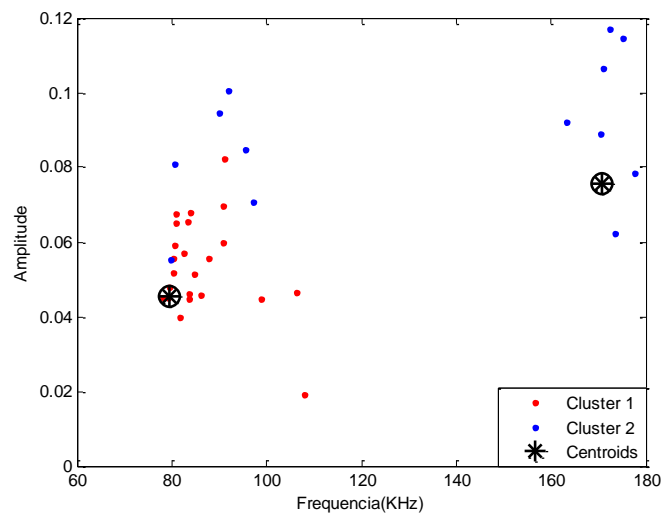


Figura 33. Classificação com 2 Classes para o Caso 2.

Tabela 8. Índice de Acerto para o Caso 2 com 2 Classes.

Ítem	Classe	Ondas PE	Ondas PI
1	Cluster 1	17	6
2	Cluster 2	1	12
Índice de Acerto		80,56%	

Como pode ser observado, tem-se um *cluster* com 23 elementos e um outro *cluster* com 13 elementos. O *Cluster 1* será subdivido em outras duas classes devido à presença de 6 ondas PI, enquanto o *Cluster 2* será mantido com todos seus elementos e seu centroide.

Para a nova subclassificação do *Cluster 1*, serão considerados os centroides produto da média das 17 ondas de PE e da média das 6 ondas de PI. Seguem abaixo os resultados obtidos da nova classificação:

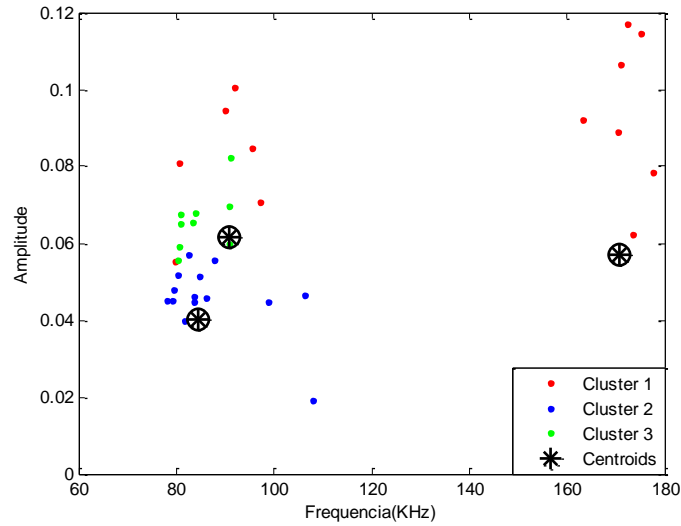


Figura 34. Classificação com 3 Classes para o Caso 2.

Tabela 9. Índice de Acerto para o Caso 2 com 3 Classes.

Item	Classe	Ondas PE	Ondas PI
1	<i>Cluster 1</i>	10	4
2	<i>Cluster 2</i>	7	2
3	<i>Cluster 3</i>	1	12
Índice de Acerto		80,56%	

O índice de acerto apresentado na Tabela 9 não teve melhoria em comparação com a classificação com 2 classes. Desta forma, na seguinte subdivisão serão considerados tanto o *Cluster 1* como o *Cluster 2*. Para isso, estimaram-se os novos centroides conforme os casos anteriores.

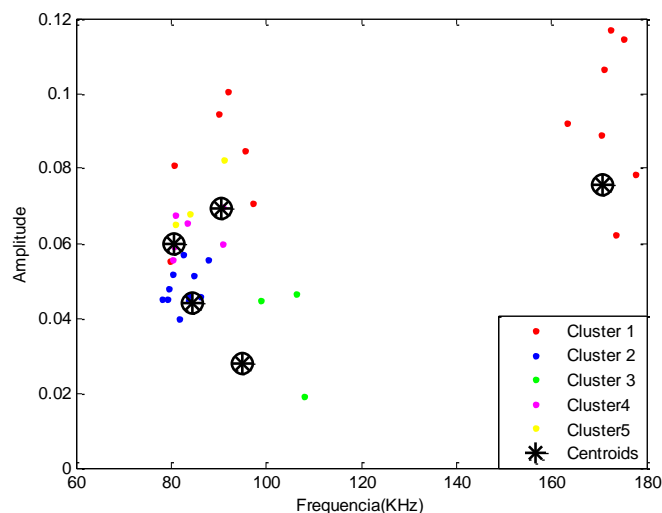


Figura 35. Classificação com 5 Classes para o Caso 2.

Tabela 10. Índice de Acerto para o Caso 2 com 5 Classes.

Item	Classe	Ondas PE	Ondas PI
1	Cluster 1	9	2
2	Cluster 2	1	2
3	Cluster 3	6	0
4	Cluster 4	1	2
5	Cluster 5	1	12
Índice de Acerto		86,11%	

Da Tabela 10, percebe-se uma melhoria no índice de acerto, o qual já pode ser considerado aceitável em comparação com os resultados no caso 1. Consequentemente, a classificação foi interrompida para não reduzir a generalização do classificador.

Neste caso, será considerado que o *Cluster 1* e o *Cluster 3* contêm as ondas de PE, enquanto que o *Cluster 2*, o *Cluster 4* e o *Cluster 5* contêm as ondas de PI.

4.2.3 Caso 3: Classificação Árvore com reclassificação

Para o terceiro caso, as considerações iniciais são as mesmas que para o Caso 1, quer dizer, que utilizaram-se a distância euclidiana e a inicialização aleatória dos centroides. Portanto, os resultados da classificação em duas classes serão os mesmos que para o caso 1, conforme à Figura 30 e à Tabela 5.

Como este caso é considerando a reclassificação, ao invés de manter os elementos e o centroide do *Cluster 1*, só manter-se-á o centroide e a nova classificação considerará esse primeiro centroide e os dois centroides obtidos de calcular a média das 18 ondas PE e as 7 ondas PI do *Cluster 2*.

Os resultados da nova classificação com as três classes definidas pelos 3 centroides estimados anteriormente mostram-se a seguir:

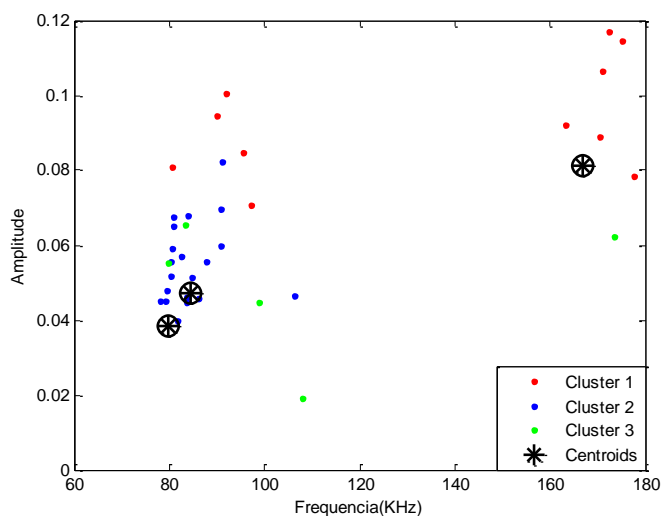


Figura 36. Classificação com 3 Classes para o Caso 3.

Tabela 11. Índice de Acerto para o Caso 3 com 3 Classes.

Item	Classe	Ondas PE	Ondas PI
1	<i>Cluster 1</i>	1	10
2	<i>Cluster 2</i>	16	4
3	<i>Cluster 3</i>	1	4
Índice de Acerto		83,33%	

Da Tabela 11, observa-se que os *clusters* 1 e 3 são predominantemente PI e que já não podem ser mais subdivididos, dado que só tem um elemento de PE. Porém, o *Cluster 2* ainda pode ser dividido em duas novas classes.

Para esta nova classificação serão calculados os dois centroides dos elementos de PE e de PI do *Cluster 2* e manter-se-ão os centroides dos *clusters* 1 e 3. Com esses 4 novos centroides será inicializado o *K-means* para quatro classes, cujos resultado apresentam-se a seguir:

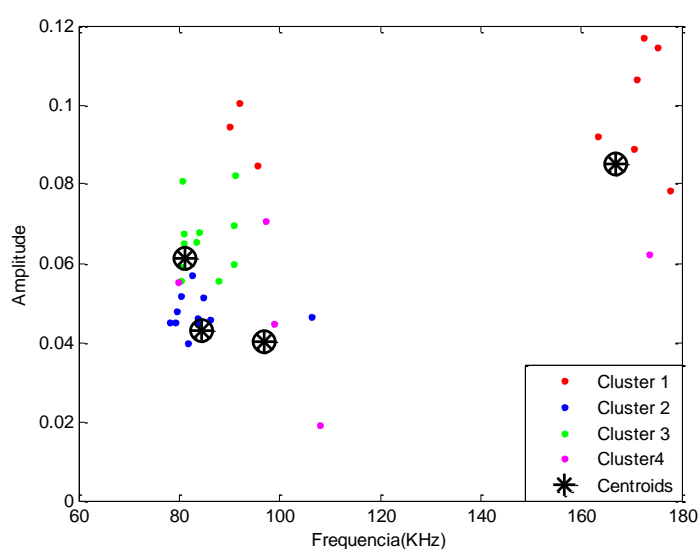


Figura 37. Classificação com 4 Classes para o Caso 3.

Tabela 12. Índice de Acerto para o Caso 3 com 4 Classes.

Item	Classe	Ondas PE	Ondas PI
1	<i>Cluster 1</i>	0	9
2	<i>Cluster 2</i>	10	1
3	<i>Cluster 3</i>	8	3
4	<i>Cluster 4</i>	0	5
Índice de Acerto		88,89%	

O índice de acerto pode ser considerado aceitável e para evitar reduzir a capacidade de generalizar do classificador a classificação foi interrompida nesse ponto. Como resultado considerar-se-ão o *Cluster 1* e o *Cluster 4* como ondas PE, e o *Cluster 2* e 3 como ondas PI.

4.2.4 Caso 4: Classificação Árvore com reclassificação e inicializando os centroides

Neste caso, as considerações iniciais são as mesmas que para o caso 2, portanto, os resultados para uma classificação em duas classes terão os mesmos resultados. A partir dos quais será feita uma nova classificação mantendo o centroide do *Cluster 2* com 13 elementos.

Para o *Cluster 1* ser dividido, serão estimados os centroides para os elementos de PE e de PI deste *cluster*. Os resultados que se obtiveram foram plotados na figura seguinte:

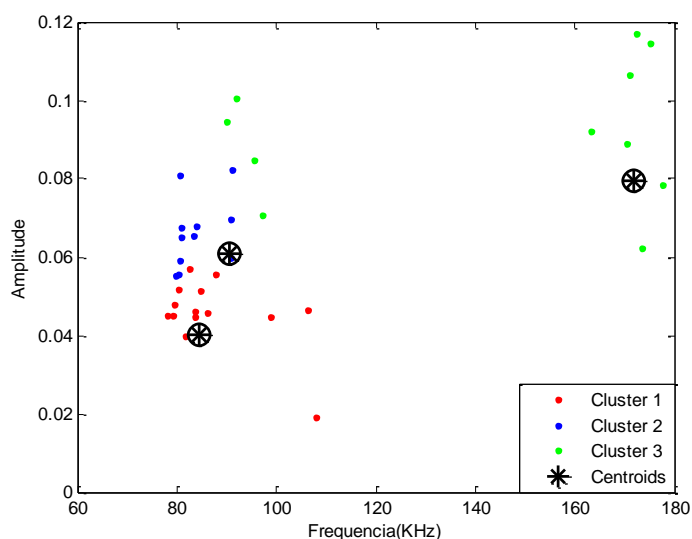


Figura 38. Classificação com 3 Classes para o Caso 4.

Tabela 13. Índice de Acerto para o Caso 4 com 3 Classes.

Item	Classe	Ondas PE	Ondas PI
1	<i>Cluster 1</i>	10	4
2	<i>Cluster 2</i>	8	3
3	<i>Cluster 3</i>	0	11
Índice de Acerto		80,56%	

Nesta nova reclassificação, observa-se que o *Cluster 3*, que anteriormente era o *Cluster 2*, agora só está composto por 11 elementos. Isso indica que dois elementos, um PE e um PI, foram reclassificados no *Cluster 1* ou no *Cluster 2*.

Dado que os *Cluster* 1 e 3 são predominantemente PE, mas ainda apresentam elementos PI, continuou-se a subdividir em duas novas classes cada um deles. Para isso, estimar-se-ão os novos centroides como a média de cada um dos elementos de PE e de PI em cada *cluster* inicial.

O *K-means* foi inicializado com os 5 novos centroides e cujos resultados são apresentado a continuação:

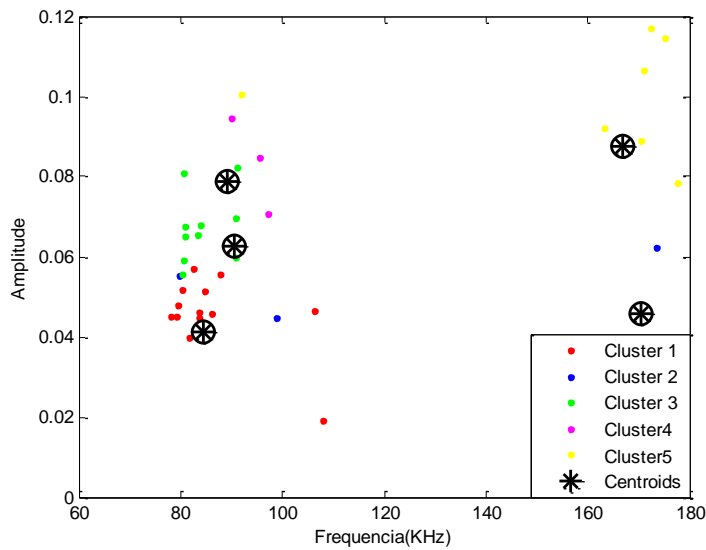


Figura 39. Classificação com 5 Classes para o Caso 4.

Tabela 14. Índice de Acerto para o Caso 4 com 5 Classes.

Item	Classe	Ondas PE	Ondas PI
1	<i>Cluster</i> 1	10	3
2	<i>Cluster</i> 2	0	3
3	<i>Cluster</i> 3	8	2
4	<i>Cluster</i> 4	0	3
5	<i>Cluster</i> 5	0	7
Índice de Acerto		86,11%	

Da Tabela 14, nota-se que o *Cluster* 1 e o *Cluster* 2 (*Cluster* 3 desta tabela) da classificação anterior, perderam um elemento cada um; enquanto o *Cluster* 3 (*Cluster* 5 nesta classificação) perdeu 4 elementos.

Apesar destas diferenças, o índice de acerto melhorou e atingiu o mesmo valor que o caso 2 para 5 classes. Destarte, é possível parar a classificação aqui para manter a capacidade de generalização do classificador.

Desta forma, ficará definido que as ondas de PE correspondem aos *clusters* 1 e 3, enquanto que as ondas de PI correspondem aos *clusters* 2, 4 e 5.

As definições dos *clusters* que correspondem às ondas de PE e de PI serão utilizadas na seção 4.5 para avaliar a acurácia destes classificadores com um conjunto maior de dados.

4.3 Classificação usando um conjunto de 68 dados

Igualmente à seção anterior, a escolha dos dados foi baseada no gráfico de pressão e deformação da trinca do capítulo 3. Só que desta vez foram considerados as ondas que se encontram na faixa de transição entre a propagação estável e a propagação instável.

O conjunto de dados está composto por 34 sinais de ondas de PE e 34 de ondas de PI, onde as ondas de PE foram adquiridas entre 4700 e 8000 segundos. Enquanto que as ondas PI foram adquiridas entre 11800 e 12300 segundos.

Na figura 40, apresentam-se a forma onda de um sinal de EA da trinca com propagação estável e a sua respectiva transformada de Fourier, e na Figura 41 a da trinca com propagação instável.

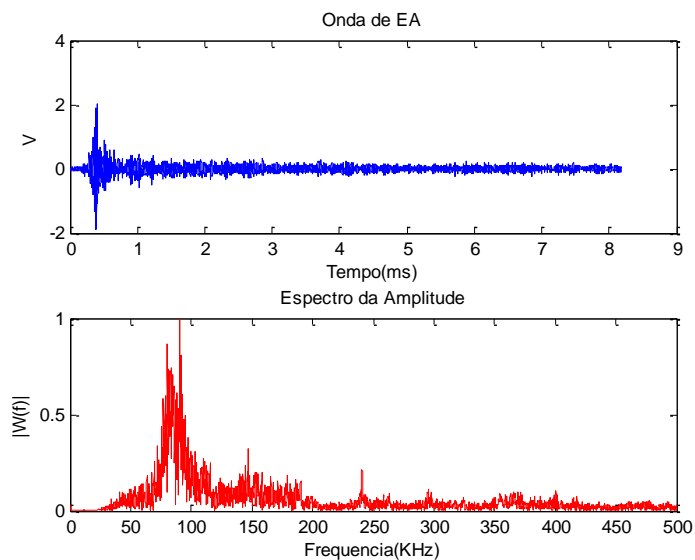


Figura 40. Forma de Onda PE.

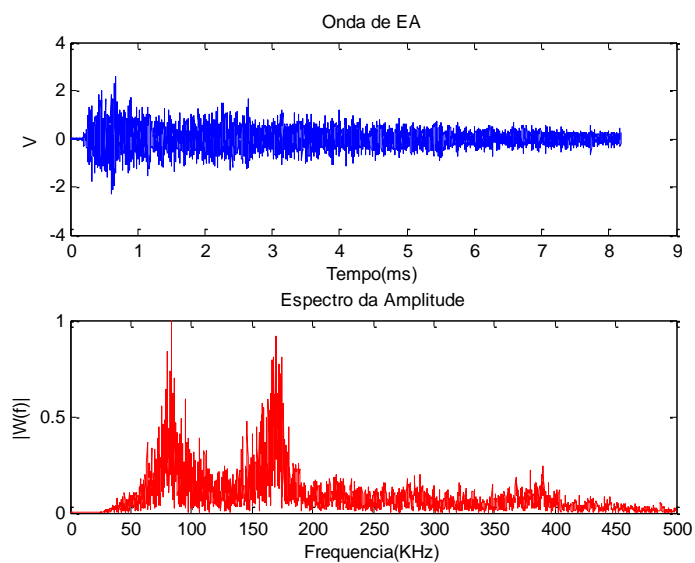


Figura 41. Forma de Onda PI.

Os casos considerados nesta seção partem dos mesmos princípios que foram definidos para a seção 4.2, assim como se mantem os critérios para deter a subdivisão de classes.

4.3.1 Caso 5: Classificação Árvore sem reclassificação

Neste caso, a matriz de entrada do classificador será de dimensão 68x4097, contendo as transformadas de Fourier das ondas de EA. E da mesma forma que na utilização dos 36 dados, para plotar os elementos de classe foram utilizados como parâmetros a frequência da maior amplitude e a média das amplitudes no espectro da onda.

A seguir apresentam-se as ondas de PE e de PI separadas corretamente:

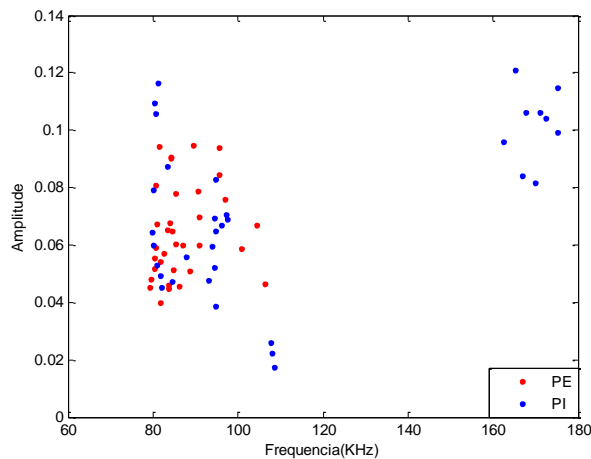


Figura 42. Ondas PE e PI.

Na Figura 42, observa-se que neste conjunto de dados existem componentes de PE e componentes de PI bastante misturados entre amplitudes que variam entre 0.04 e 0.08 e entre as frequências de 80 e 100 KHz.

Neste caso, foram utilizados como parâmetros para o *K-means* a distância euclidiana entre a entrada e os centroides, a inicialização aleatória dos centroides e uma primeira classificação de duas classes. Os resultados obtidos foram os seguintes:

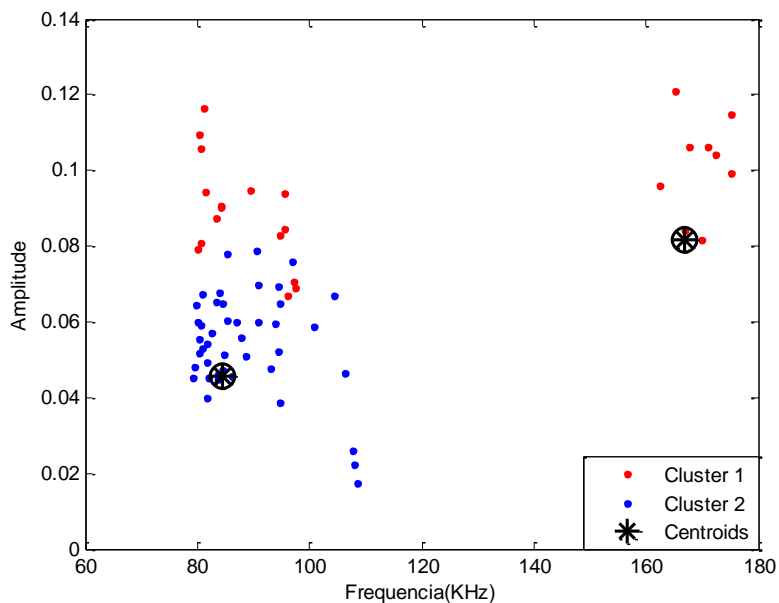


Figura 43. Classificação com duas Classes para o Caso 5.

Tabela 15. Índice de acerto para o Caso 5 com 2 Classes

Item	Classe	Ondas PE	Ondas PI
1	Cluster 1	7	18
2	Cluster 2	27	16
Índice de Acerto		66,18%	

Da Tabela 15, observa-se que o índice de acerto é baixo e que o erro das classes em ambas as classes também não é o adequado. Portanto, os dois *clusters* serão subdivididos em duas classes mais.

Para realizar a subdivisão, foram estimados os centroides dos elementos de PE e de PI do *Cluster 1*, com os quais inicializou-se o classificador e para obter duas novas classes. Procedeu-se da mesma forma com o *Cluster 2*. Desta forma, obtiveram-se os resultados apresentados abaixo:

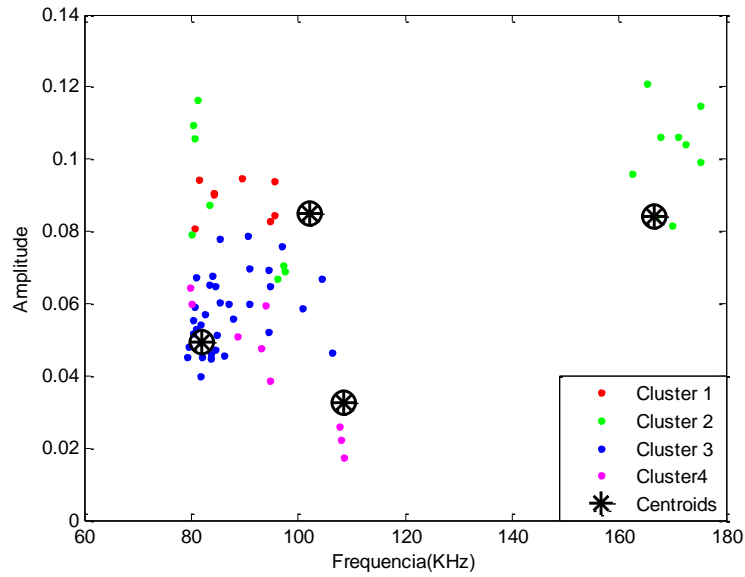


Figura 44. Classificação com 4 Classes para o Caso 5.

Tabela 16. Índice de Acerto para o Caso 5 com 4 Classes.

Item	Classe	Ondas PE	Ondas PI
1	<i>Cluster 1</i>	7	1
2	<i>Cluster 2</i>	0	17
3	<i>Cluster 3</i>	28	6
4	<i>Cluster 4</i>	1	8
Índice de Acerto		88,24%	

Destes resultados observa-se um incremento no índice de acerto e que os erros dos *clusters* 1, 2 e 4 são mínimos, por esse motivo, eles não serão mais subdivididos. Porém, no *Cluster 3* ainda existem elementos de PI embora a predominância seja de PE.

Baseado nessa análise, procedeu-se a dividir em outras duas classes o *Cluster 3*, tomando em consideração a média dos elementos do *cluster*. Os resultados que se obtiveram são apresentados na Figura 45 e na Tabela 17.

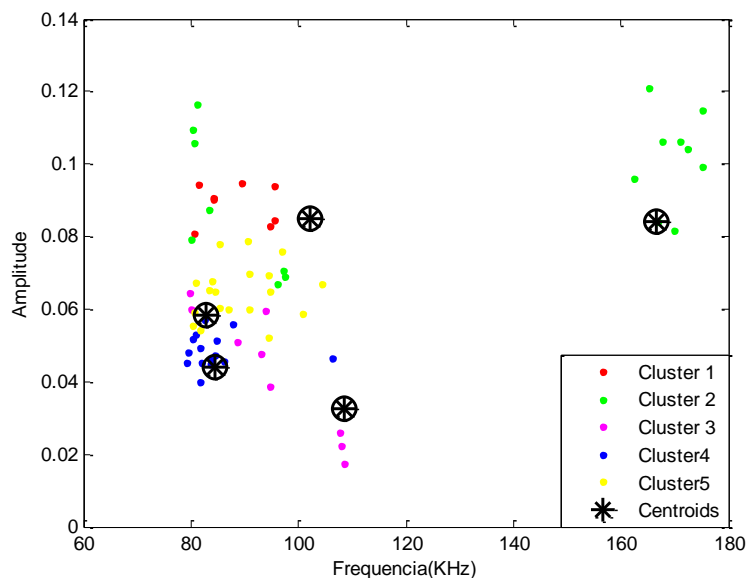


Figura 45. Classificação com 6 Classes para o Caso 5.

Tabela 17. Índice de Acerto para o Caso 5 com 5 Classes.

Item	Classe	Ondas PE	Ondas PI
1	Cluster 1	7	1
2	Cluster 2	0	17
3	Cluster 3	12	3
4	Cluster 4	16	3
5	Cluster 5	1	8
Índice de Acerto		88,24%	

Devido a que o índice de acerto não melhorou com esta nova classificação e para não reduzir a capacidade de generalização do classificador optou-se por parar a classificação nesta fase. Desta forma, tomaram-se como referência os *Clusters* 1, 3 e 4 para classificar as ondas PE; enquanto que os *Clusters* 2 e 5 para as ondas de PI.

4.3.2 Caso 6: Classificação árvore sem reclassificação e inicializando os centroides

Neste caso, tiveram-se as mesmas considerações que no Caso 5, diferenciando-os na inicialização dos centroides. Enquanto no Caso 5 a inicialização foi aleatória, aqui no Caso 6, os centroides com que se inicializou o *K-means* foram obtidos mediante a média dos elementos de PE e de PI do conjunto total de dados.

Da primeira classificação em duas classes com as considerações anteriores, obtiveram-se os seguintes resultados:

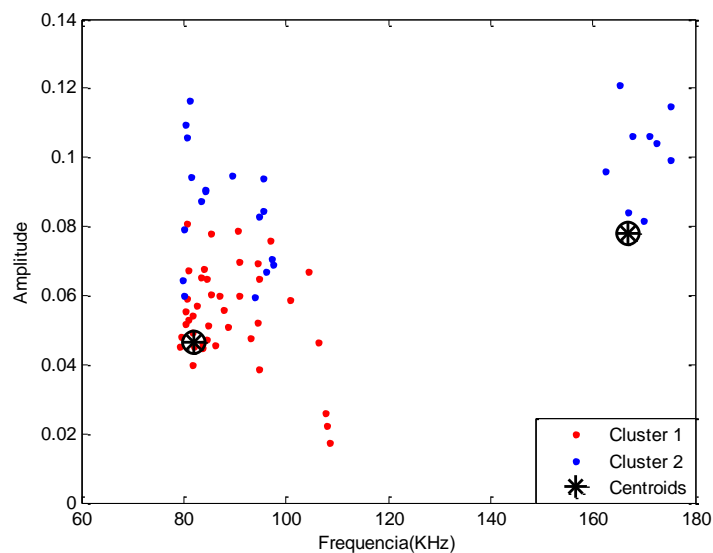


Figura 46. Classificação com 2 Classes para o Caso 6.

Tabela 18. Índice de Acerto para o Caso 6 com 2 Classes.

Item	Classe	Ondas PE	Ondas PI
1	Cluster 1	28	13
2	Cluster 2	6	21
Índice de Acerto		72,06%	

Da Tabela 18 nota-se que os *Clusters* não têm uma predominância, seja dos elementos de PE ou dos elementos de PI. Além disso, o índice de acerto ainda pode ser melhorado. Portanto, cada *cluster* será dividido em duas novas classes. As 4 classes resultantes desta divisão são as seguintes:

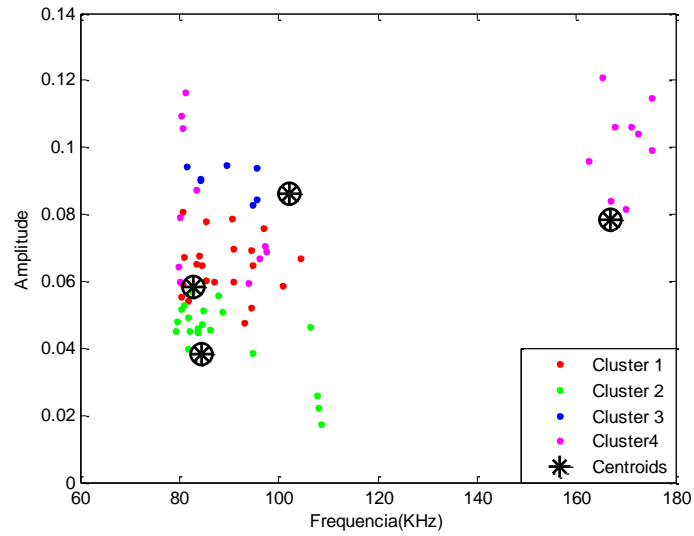


Figura 47. Classificação com 4 Classes para o Caso 6.

Tabela 19. Índice de Acerto para o Caso 6 com 4 Classes.

Item	Classe	Ondas PE	Ondas PI
1	Cluster 1	17	4
2	Cluster 2	11	9
3	Cluster 3	6	1
4	Cluster 4	0	20
Índice de Acerto		79,41%	

Os resultados tiveram uma leve melhoria, ficando bem definidos os *clusters* 3 e 4. Porém, os *clusters* 1 e 2 ainda apresentam ondas de PI, para o qual eles foram divididos em duas novas classes, considerando a média dos seus componentes como centroides. Os resultados obtidos apresentam-se a continuação:

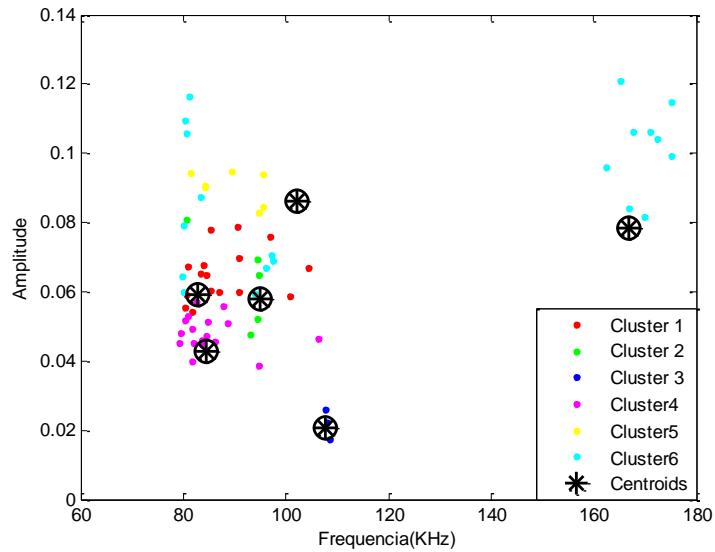


Figura 48. Classificação com 6 Classes para o Caso 6.

Tabela 20. Índice de Acerto para o Caso 6 com 6 Classes.

Item	Classe	Ondas PE	Ondas PI
1	Cluster 1	16	0
2	Cluster 2	1	4
3	Cluster 3	11	6
4	Cluster 4	0	3
5	Cluster 5	6	1
6	Cluster 6	0	20
Índice de Acerto		88,24%	

Destes resultados, percebe-se que os *clusters* 1, 2, 4, 5 e 6 estão bem definidos; porém o *Cluster* 3 ainda apresenta alguns intrusos. Este último foi dividido novamente para tentar aumentar o índice de acerto do classificador.

Os resultados de classificação com 7 classes obtidos desta vez são apresentados na forma seguinte:

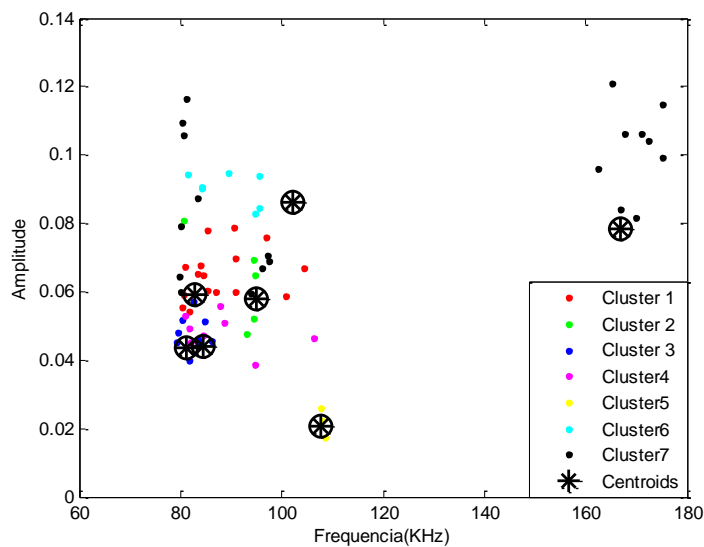


Figura 49. Classificação com 7 Classes para o Caso 6.

Tabela 21. Índice de Acerto para o Caso 6 com 7 Classes.

Ítem	Classe	Ondas PE	Ondas PI
1	Cluster 1	16	0
2	Cluster 2	1	4
3	Cluster 3	9	0
4	Cluster 4	2	6
5	Cluster 5	0	3
6	Cluster 6	6	1
7	Cluster 7	0	20
Índice de Acerto		94,12%	

Nestes resultados, já se conseguiram classes melhor definidas e o índice de acerto é bastante positivo. Conseqüentemente, já não foi necessário fazer mais subclassificações e foram definidos os *cluster* 1, 3 e 6 como os pertencentes às ondas de PE; e os *clusters* 2, 4, 5 e 7 como os pertencentes às ondas de PI.

4.3.3 Caso 7: Classificação Árvore com reclassificação

Como as considerações deste caso são as mesmas que se tiveram no caso 5, a primeira classificação em duas classes será também a mesma. A diferença de enfoque entre ambos os casos radica que enquanto no caso 5 as classes bem definidas eram man-

tidas e só eram subdividas aquelas que não, neste caso só se mantiveram os centroides das classes bem definidas e para a nova classificação foram usados estes centroides e aqueles estimados com as médias dos elementos de PE e de PI daquelas classes que ainda podiam ser divididos.

Embora que as considerações tenham sido as mesmas, os resultados não foram os mesmos. Neste caso, o classificador convergiu na classificação de duas classes, uma com 39 elementos e a segunda com 29. Estes resultados são apresentados na figura seguinte:

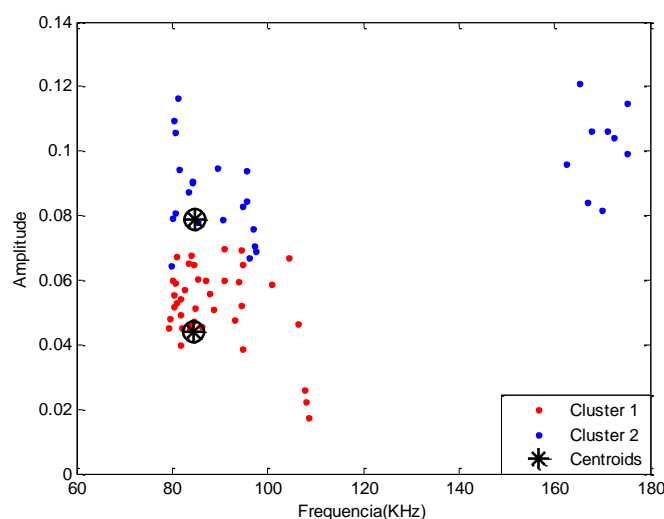


Figura 50. Classificação com 2 Classes para o Caso 7.

Tabela 22. Índice de Acerto para o Caso 7 com 2 Classes.

Ítem	Classe	Ondas PE	Ondas PI
1	Cluster 1	24	15
2	Cluster 2	10	19
Índice de Acerto		63,24%	

Como pode ser observado na Tabela 22 o índice de acerto do classificador está baixo, portanto, ainda devem ser subdivididos cada um dos *clusters*. Para esta subdivisão foram estimados os centroides como nos casos anteriores, mas desta vez a classificação foi feita com o conjunto total de dados. Os resultados foram os seguintes:

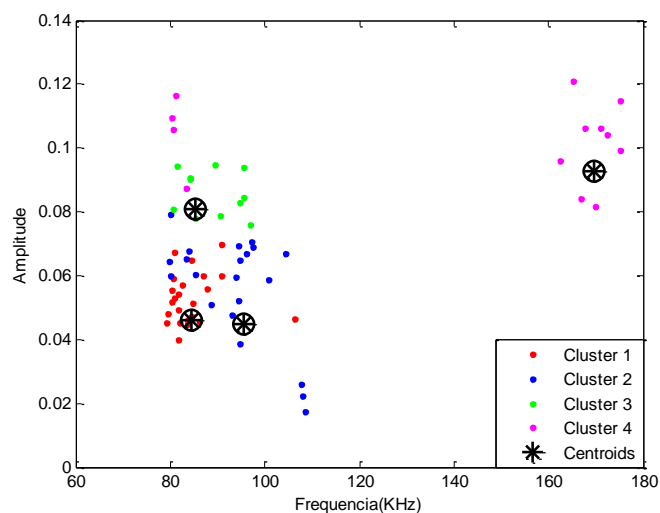


Figura 51. Classificação com 4 Classes para o Caso 7.

Tabela 23. Índice de Acerto para o Caso 7 com 4 Classes.

Item	Classe	Ondas PE	Ondas PI
1	Cluster 1	18	5
2	Cluster 2	6	15
3	Cluster 3	10	1
4	Cluster 4	0	13
Índice de Acerto		82,35%	

Da Tabela 23 observa-se que o índice de acerto aumentou consideravelmente e que os *clusters* 3 e 4 estão bem definidos, o primeiro como PE e o segundo como PI. Desta forma, só foi necessário dividir os *clusters* 1 e 2.

A subdivisão foi realizada calculando a média dos elementos de PE e de PI dos *clusters* 1 e 2, e mantendo os centroides finais dos *clusters* 3 e 4. Os resultados da nova classificação são apresentados na Figura 52 e na Tabela 24.

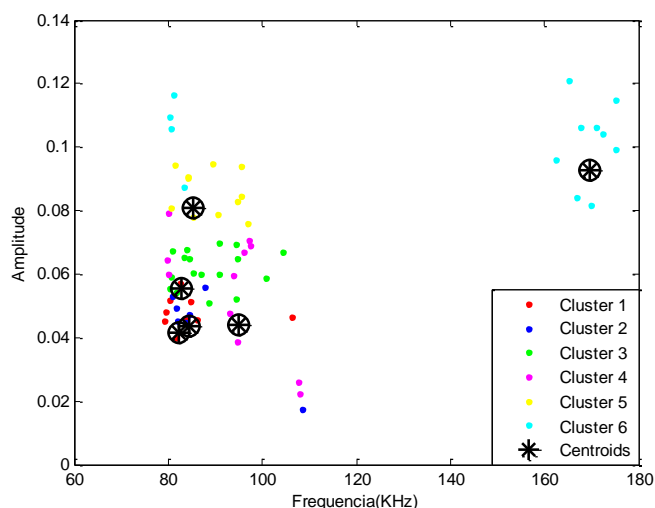


Figura 52. Classificação com 6 Classes para o Caso 7.

Tabela 24. Índice de Acerto para o Caso 7 com 6 Classes.

Ítem	Classe	Ondas PE	Ondas PI
1	Cluster 1	9	0
2	Cluster 2	1	6
3	Cluster 3	14	3
4	Cluster 4	0	11
5	Cluster 5	10	1
6	Cluster 6	0	13
Índice de Acerto		92,65%	

Como pode ser observado na tabela anterior, o índice de acerto é alto e todas as *clusters* encontram-se bem definidas. Portanto, para não prejudicar a capacidade de generalização do classificador a divisão em subclasses parou-se aqui. Com isso ficam definidas as seguintes classificações: *Cluster 1*, *Cluster 3* e *Cluster 5* são para as ondas PE; enquanto o *Cluster 2*, o *Cluster 4* e o *Cluster 6* são para as ondas PI.

4.3.4 Caso 8: Classificação Árvore com reclassificação e inicializando os centroides

Neste caso, foram tomadas em conta as mesmas considerações que no caso 6, com a diferença de que o conjunto total de dados é reclassificado. Para a reclassificação procedeu-se da mesma forma que no caso 7.

A primeira classificação em duas classes teve os mesmos resultados que no caso 6, devido a que a inicialização foi a mesma. Assim, baseado nos resultados apresentados na Figura 46 e na Tabela 18, procedeu-se a estimar os centroides dos elementos PE e PI de cada um dos *clusters*.

Com os centroides estimados, reclassificaram-se o conjunto total de dados e obtiveram-se os seguintes resultados:

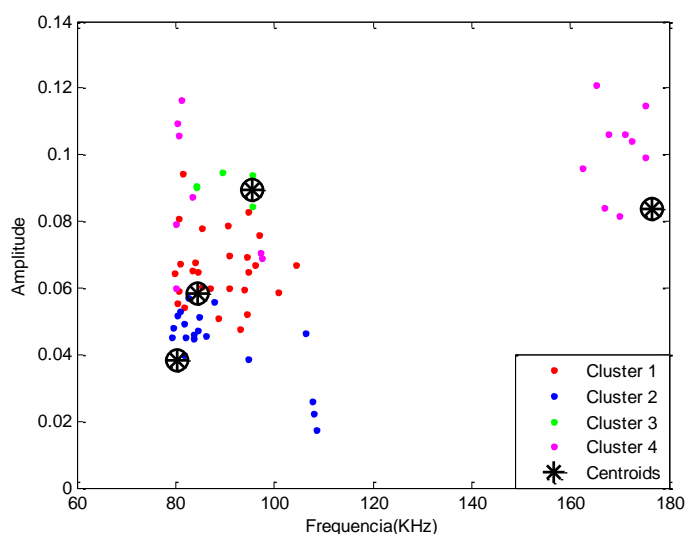


Figura 53. Classificação com 4 Classes para o Caso 8.

Tabela 25. Índice de Acerto para o Caso 8 com 4 Classes.

Ítem	Classe	Ondas PE	Ondas PI
1	Cluster 1	19	8
2	Cluster 2	10	9
3	Cluster 3	5	0
4	Cluster 4	0	17
Índice de Acerto		73,53%	

Na Tabela 25, observam-se que os *Cluster 1* e *2* contêm elementos PE predominantes, mas também têm elementos PI, desta forma, ainda foi necessário dividir cada um deles. Já no caso dos *Clusters 3* e *4* se vê que são predominantemente PE e PI, respectivamente.

Portanto, mantiveram-se os centroides dos *Clusters 3* e *4*; e estimaram-se dois centroides novos para o *Cluster 1* e dois para o *Cluster 2*, conforme os casos anteriores. A nova reclassificação tem a seguinte agrupação:

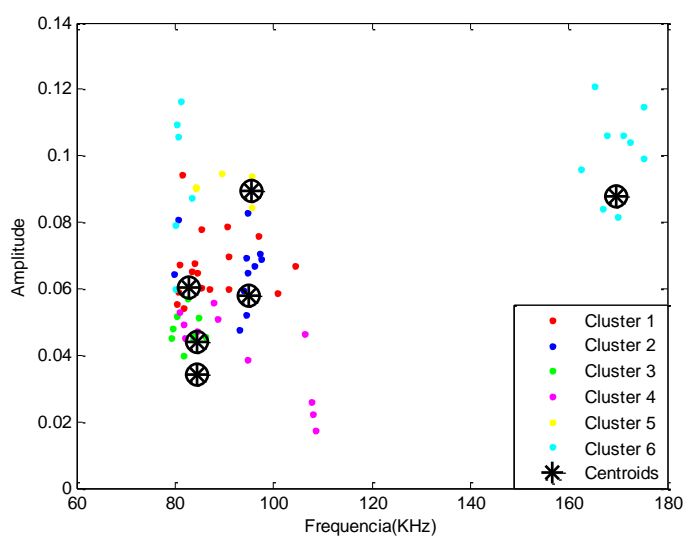


Figura 54. Classificação com 6 Classes para o Caso 8.

Tabela 26. Índice de Acerto para o Caso 8 com 6 Classes.

Item	Classe	Ondas PE	Ondas PI
1	<i>Cluster 1</i>	17	0
2	<i>Cluster 2</i>	1	10
3	<i>Cluster 3</i>	9	0
4	<i>Cluster 4</i>	2	9
5	<i>Cluster 5</i>	5	0
6	<i>Cluster 6</i>	0	15
Índice de Acerto		95,88%	

Como pode ser observado na Tabela 26, o índice de acerto é alto e a nova classificação apresenta classes bem definidas, consequentemente, já não é necessário fazer novas classificações.

As ondas de PE serão consideradas aquelas que pertençam aos *clusters* 1, 2 e 3; e as ondas aquelas que pertençam aos *clusters* 2, 4 e 6.

4.4 Validação Cruzada

Para verificar a acurácia dos classificadores desenvolvidos nas seções anteriores, optou-se por usar o método de *validação cruzada*, que consiste em dividir o conjunto de dados em k partições, das quais uma será usada como conjunto de teste e as $k-1$ partições restantes são utilizadas para o treinamento. Este método é iterativo porque será realizada k vezes para que cada partição seja em algum momento o conjunto de teste.

Devido à pouca quantidade de dados utilizados na seção 4.2, só foi aplicada a validação cruzada ao conjunto de 68 dados. Para isso, foram consideradas 3 partições com um total de 22 elementos cada uma.

Para realizar essa divisão previamente se eliminou um elemento de PE e um elemento de PI, restando um total de 66 elementos. Cada partição foi denominada como A, B e C; e a escolha de cada elemento pertencente a cada partição foi de tal forma que ficasse espalhada no tempo e que contivesse tanto ondas de PE como de PI.

A cada 3 elementos um era assignado a uma partição, sendo que o primeiro foi escolhido para ser da partição A, o segundo para ser da partição B e o terceiro para ser da partição C. E assim sucessivamente até completar os 66 elementos.

Os casos considerados para aplicação da validação cruzada foram os casos 7 e 8 que usam a reclassificação e que finalizaram com 6 classes.

4.4.1 Caso 1: Classificação com centroides aleatórios

A primeira fase para realizar a *Cross-Validation* foi a de tomar as partições A e B como conjuntos de treinamento e a partição C como conjunto de teste. Na Figura 55, foram plotados os conjuntos A e B.

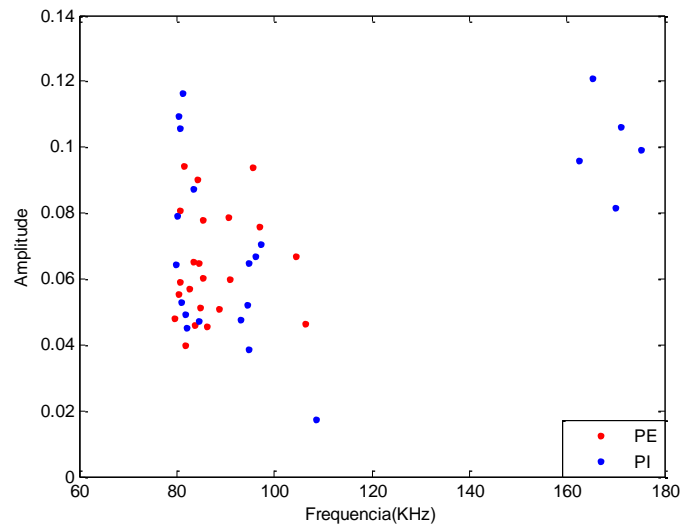


Figura 55. Ondas PE e PI das partições A e B.

O processo de treinamento das partições seguiu o mesmo padrão que nas seções 4.2 e 4.3. Destarte, as seguintes figuras mostrarão a subdivisão das classes até chegar a um total de 6 classes, onde a Figura 56 mostra a primeira classificação em duas classes, a Figura 57 em 4 classes e a Figura 58 em 6 classes.

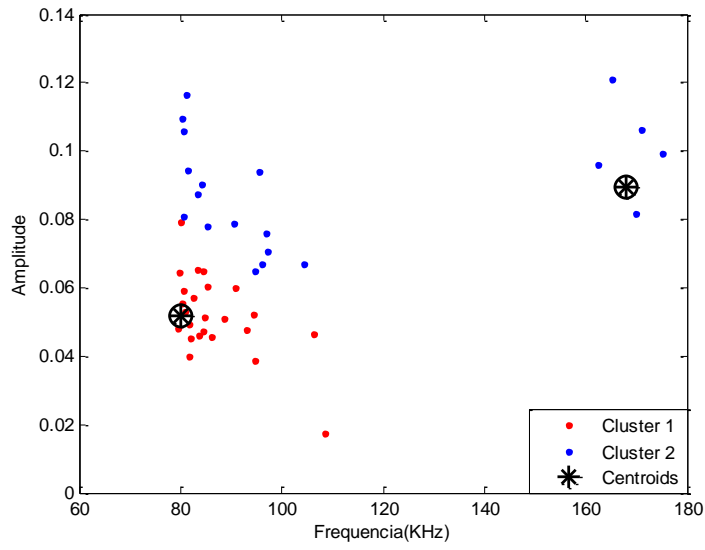


Figura 56. Classificação com 2 classes para as partições A e B.

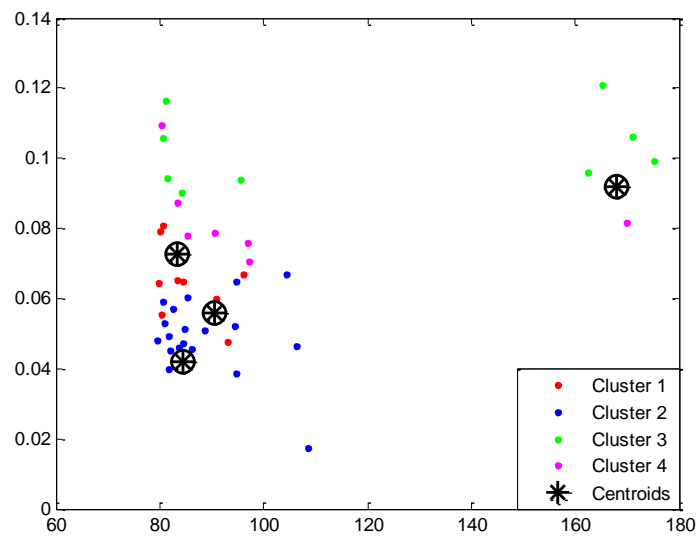


Figura 57. Classificação com 4 Classes para as partições A e B.

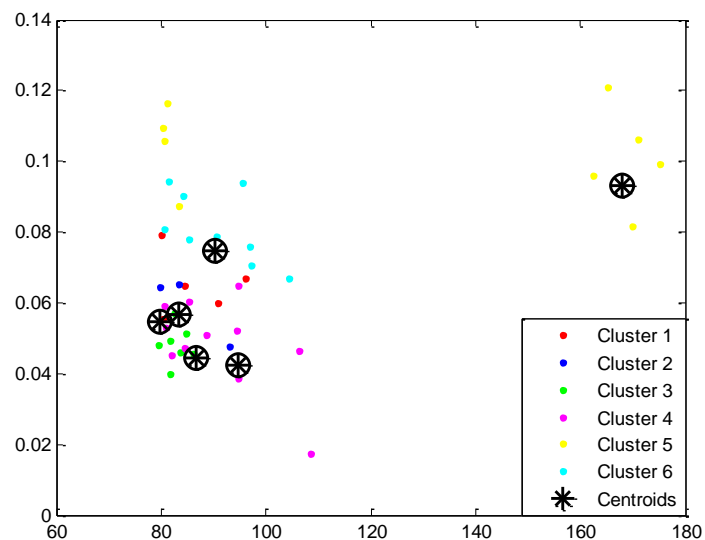


Figura 58. Classificação com 6 Classes para as partições A e B.

Da Figura 58 e da Tabela 24, infere-se que as ondas de PE são as que pertencem aos *clusters* 1, 3 e 5, enquanto que as ondas de PI pertencem aos *clusters* 2, 4 e 6. Com esta classificação, procedeu-se a validar a acurácia com a partição C, cujo resultado mostra-se na Figura 59.

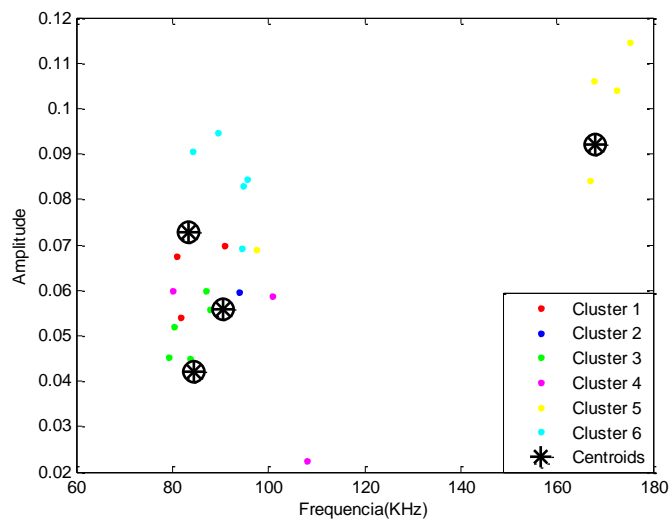


Figura 59. Teste com a partição C.

Tabela 27. Índice de Acerto com a Partição C.

Item	Classe	Ondas PE	Ondas PI
1	Cluster 1	3	0
2	Cluster 2	0	1
3	Cluster 3	4	1
4	Cluster 4	1	2
5	Cluster 5	0	5
6	Cluster 6	3	2
Índice de Acerto		54,55%	

Na segunda fase, as partições A e C foram usadas para o treinamento e a partição B para teste. Na Figura 60, foram plotados os conjuntos A e C.

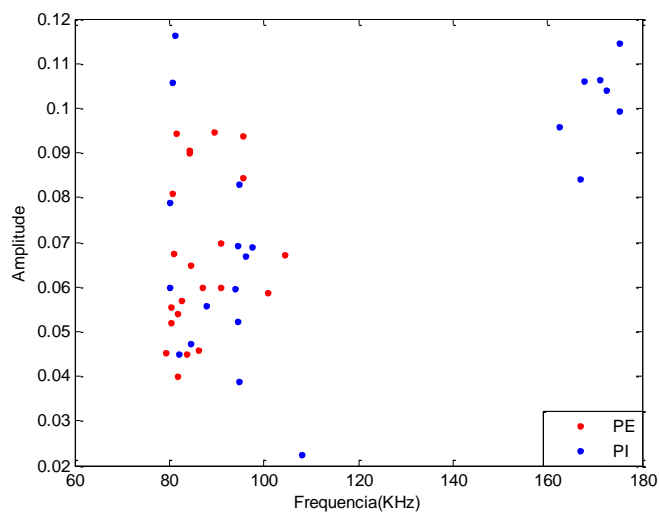


Figura 60. Formas de Onda PE e PI das partições A e C.

Após realizar o treinamento da mesma forma em que se procedeu anteriormente, a classificação com 6 classes ficou da seguinte forma:

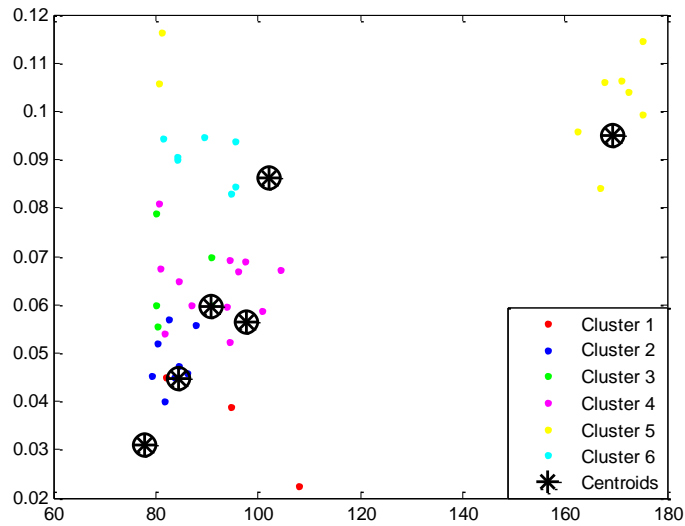


Figura 61. Classificação com 6 Classes para as partições A e C.

Para o teste com o conjunto B, os *clusters* 1, 3, e 5 foram considerados como ondas PE; e os *clusters* 2, 4 e 6 como ondas PI. Os resultados obtidos são apresentados na Figura 62 e na Tabela 28.

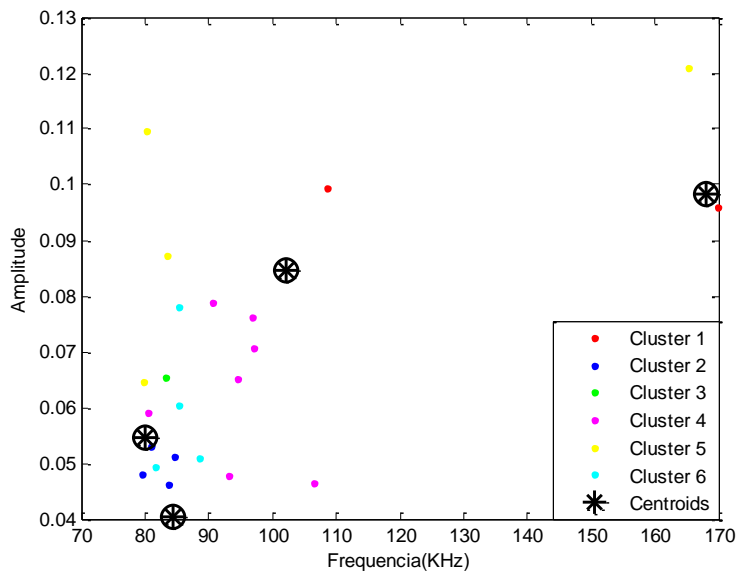


Figura 62. Teste com a partição B.

Tabela 28. Índice de Acerto da partição B.

Item	Classe	Ondas PE	Ondas PI
1	Cluster 1	0	2
2	Cluster 2	3	1
3	Cluster 3	1	0
4	Cluster 4	4	3
5	Cluster 5	0	4
6	Cluster 6	3	1
Índice de Acerto		27,27%	

Na terceira e última fase, as partições B e C foram usadas para o treinamento e a partição A para teste. Na Figura 63, foram plotados os conjuntos B e C.

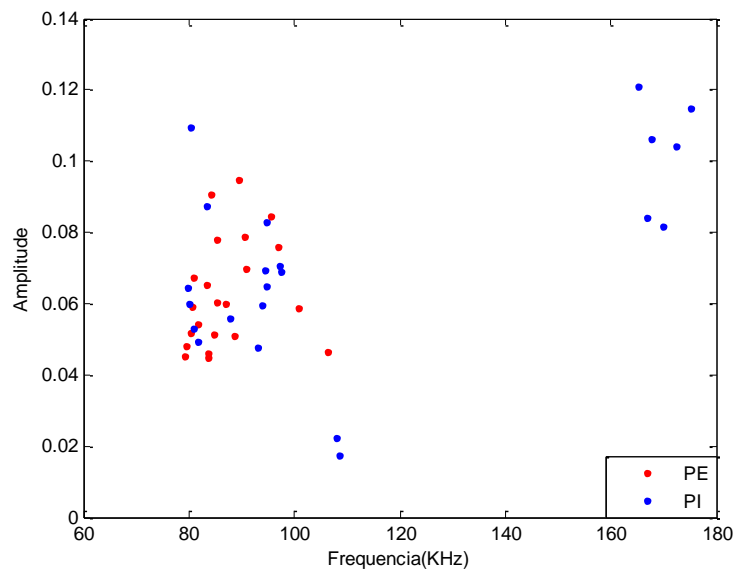


Figura 63. Formas de onda PE e PI para as partições B e C.

O treinamento foi realizado da mesma forma que nas fases anteriores e o resultado da classificação para este caso com 6 classes é apresentado na Figura 64. Onde novamente as ondas de PE correspondem aos *clusters* 1, 3 e 5; e as ondas PI aos *clusters* 2, 4 e 6.

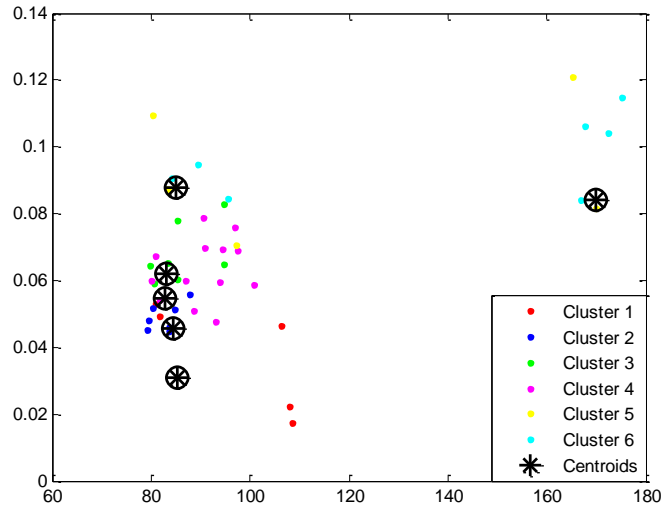


Figura 64. Classificação com 6 Classes para as partições B e C.

Com a classificação finalizada, testou-se a acurácia com a partição A, cujos resultados encontram-se detalhados na Figura 65 e na Tabela 29.

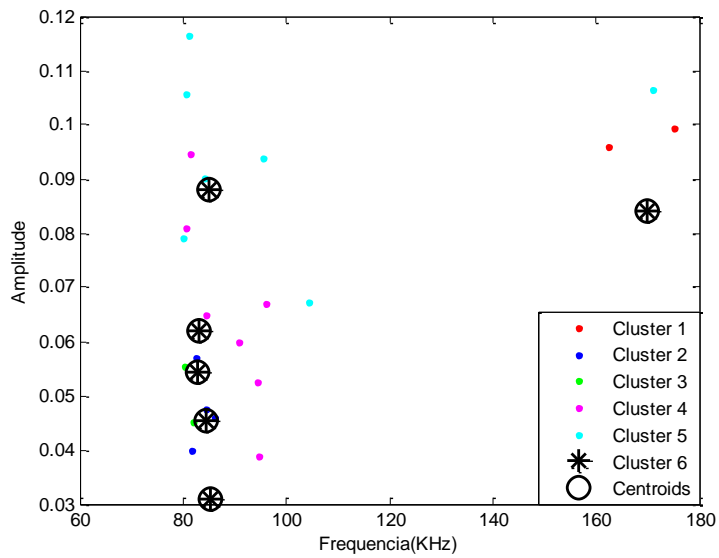


Figura 65. Teste com a partição A.

Tabela 29. Índice de Acerto com a partição A.

Item	Classe	Ondas PE	Ondas PI
1	Cluster 1	0	2
2	Cluster 2	3	1
3	Cluster 3	1	1
4	Cluster 4	4	3
5	Cluster 5	0	0
6	Cluster 6	3	4
Índice de Acerto		40,91%	

Uma vez que foram estimados os índices de acerto para cada uma das fases, estimou-se a média dos mesmos para calcular a eficiência da predição.

Tabela 30. Eficiência de Predição.

Item	Partição de Teste	Índice de Acerto
1	C	54,55
2	B	27,27
3	A	40,91
Eficiência de Predição		40,91%

4.4.2 Caso 2: Classificação com inicializando os centroides

Neste caso, foi considerado o desenvolvimento dos classificadores inicializando os centroides como no caso 8 da seção 4.3.

Na primeira fase, os conjuntos de treinamento foram A e B, e o conjunto C foi o do teste. Na figura seguinte foram plotadas as ondas de PE e de PI das partições A e B.

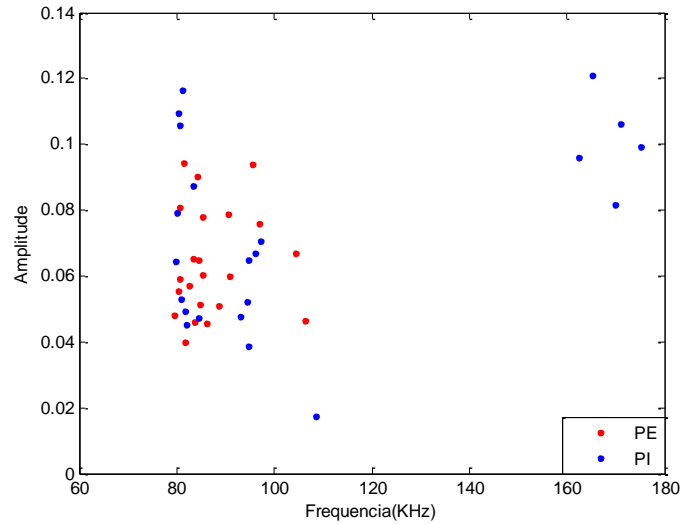


Figura 66. Formas de Onda PE e PI das partições A e B.

A classificação considerou igualmente a divisão de classes, e como foi realizado na seção 4.3 a classificação final está composta por 6 classes. Portanto, a seguir será mostrado o resultado final para esta fase.

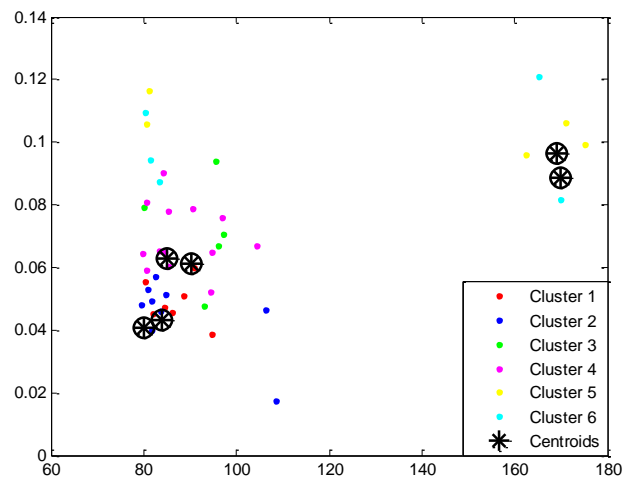


Figura 67. Classificação com 6 Classes para as partições A e B.

Para o teste com a partição C, foram considerados que os *clusters* 1, 3 e 5 são para as ondas de PE; enquanto que os *clusters* 2, 4 e 6 para as ondas PI. Os resultados foram os seguintes:

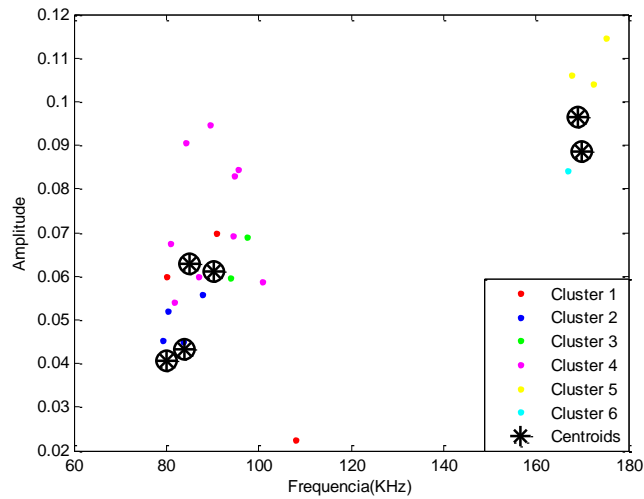


Figura 68. Teste com a partição C.

Tabela 31. Índice de Acerto para a partição C.

Item	Classe	Ondas PE	Ondas PI
1	Cluster 1	1	2
2	Cluster 2	3	1
3	Cluster 3	0	2
4	Cluster 4	7	2
5	Cluster 5	0	3
6	Cluster 6	0	1
Índice de Acerto		22,73%	

A fase 2 considera o treinamento com as partições A e C, cuja distribuição foi plotada na Figura 69. E para o teste foi utilizada a partição B.

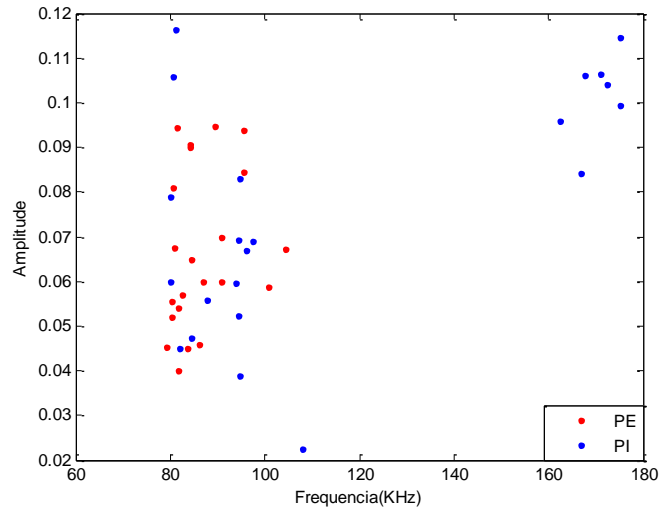


Figura 69. Formas de Onda PE e PI para as partições A e C.

Os *clusters* que definem as ondas de PE e as ondas de PI foram os mesmos que na fase anterior, como pode ser observado na Figura 70 a seguir.

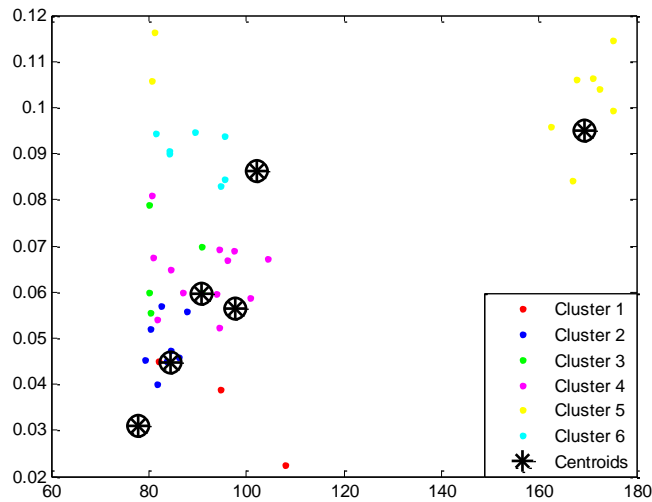


Figura 70. Classificação com 6 Classes para as partições A e C.

Ao testar esta classificação com a partição B, a classificação foi conforme a Figura 71 e a Tabela 32.

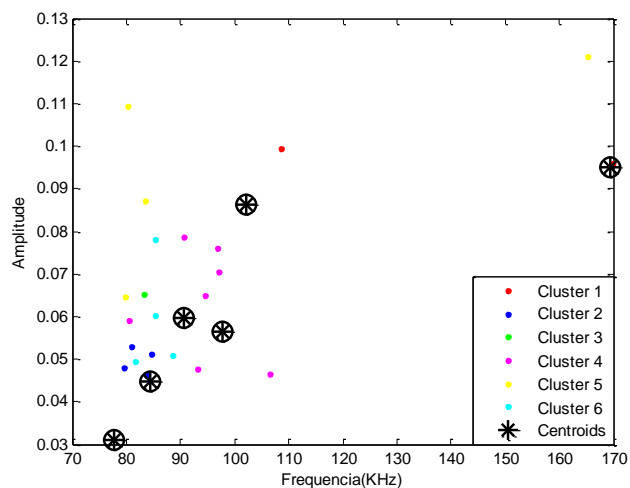


Figura 71. Teste com a partição B.

Tabela 32. Índice de Acerto para a Partição B.

Ítem	Classe	Ondas PE	Ondas PI
1	Cluster 1	0	2
2	Cluster 2	3	1
3	Cluster 3	1	0
4	Cluster 4	4	3
5	Cluster 5	0	4
6	Cluster 6	3	1
Índice de Acerto		27,27%	

Por último na fase 3, os conjuntos de treinamento foram B e C, cuja distribuição mostra-se na Figura 72, e o conjunto A foi utilizado para o teste da classificação.

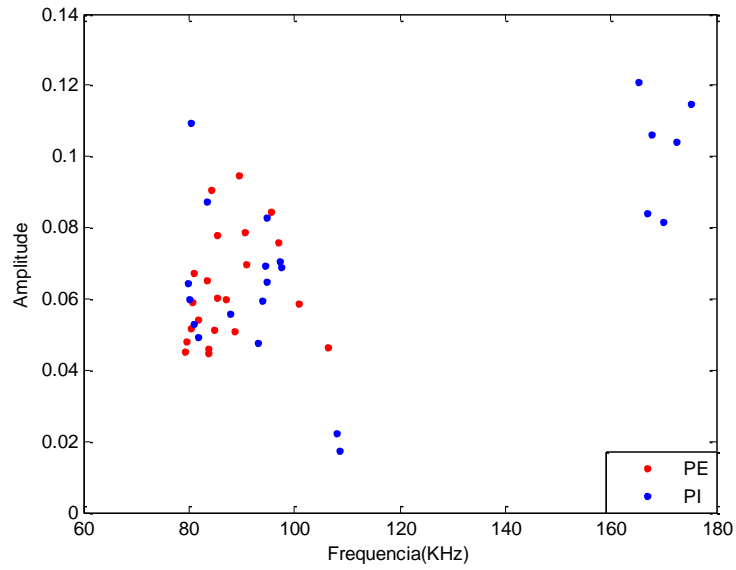


Figura 72. Formas de Onda PE e PI para as partições B e C.

A classificação com 6 classes ficou distribuída como mostra-se na seguinte Figura 73. E as ondas de PE encontram-se classificadas nos *clusters* 1, 2, 3 e 5; enquanto que as ondas PI nos *clusters* 4 e 6.

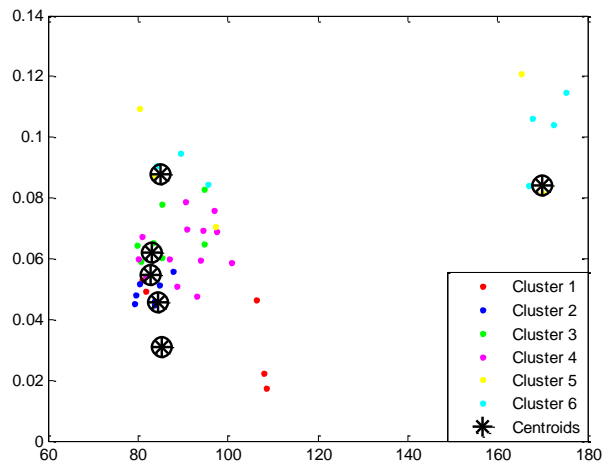


Figura 73. Classificação com 6 Classes para as partições B e C.

Ao testar esta classificação com a partição A, a classificação foi conforme a Figura 74 e a Tabela 33.

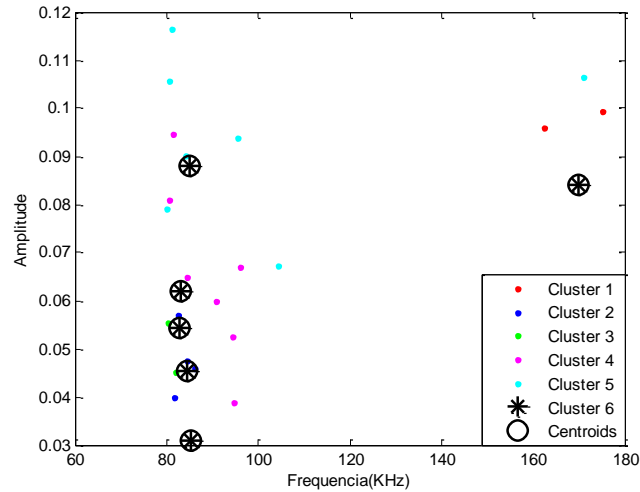


Figura 74. Teste com a partição A.

Tabela 33. Índice de Acerto com a partição A.

Item	Classe	Ondas PE	Ondas PI
1	Cluster 1	0	2
2	Cluster 2	3	1
3	Cluster 3	1	1
4	Cluster 4	4	3
5	Cluster 5	0	0
6	Cluster 6	3	4
Índice de Acerto		50,00%	

Uma vez que foram estimados os índices de acerto para cada uma das fases, estimou-se a média dos mesmos para calcular a eficiência da predição.

Tabela 34. Eficiência de Predição.

Item	Partição de Teste	Índice de Acerto
1	C	22,73
2	B	27,27
3	A	50,00
Eficiência de Predição		33,33%

4.5 Acurácia dos Classificadores desenvolvidos com 36 dados

Os resultados obtidos na seção anterior não foram satisfatórios, devido as baixas taxas de acerto, e portanto, à pouca capacidade de generalização dos classificadores desenvolvidos com as 68 formas de onda.

Conseqüentemente, faz-se necessário testar a acurácia dos classificadores desenvolvidos com as 36 formas de onda. Uma forma viável para realizar isso é a de utilizar os centroides obtidos com as melhores classificações para cada caso analisado na seção 4.2 e classificar o conjunto de 68 dados.

O caso 1 da seção 4.2 finalizou a classificação com 4 classes e com um índice de acerto de 91,66%. Ao utilizar os centroides destas 4 classes e classificar o conjunto maior de dados obtiveram-se os seguintes resultados:

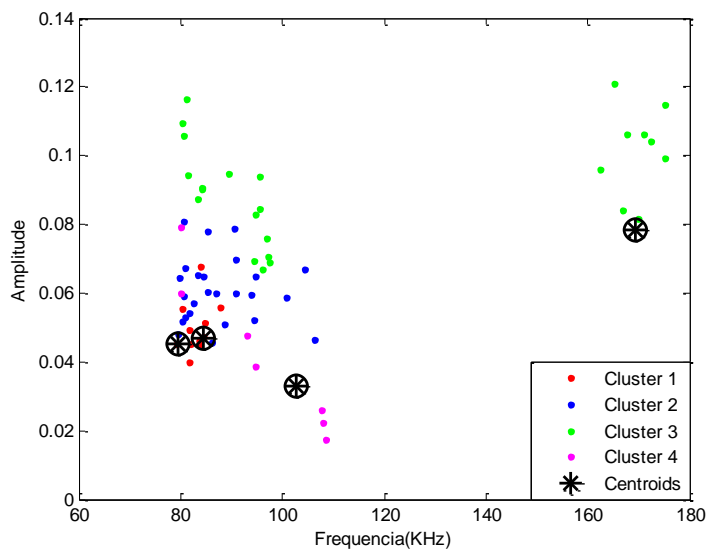


Figura 75. Classificação com os centroides do Caso 1.

Deve-se lembrar que os *clusters* 1, 2 e 4 classificam as ondas como PI, e o *cluster* 3 como PE.

Tabela 35. Índice de Acerto com os centroides do Caso 1.

Item	Classe	Ondas PE	Ondas PI
1	Cluster 1	7	4
2	Cluster 2	20	5
3	Cluster 3	7	18
4	Cluster 4	0	7
Índice de Acerto		33,82%	

O caso 2 finalizou a classificação com 5 classes e com um índice de acerto de 86,11%. Onde os *clusters* 1 e 3 definem as ondas PE; e os *clusters* 2, 4 e 5 as ondas PI.

Ao utilizar os centroides destas 5 classes e classificar o conjunto maior de dados obtiveram-se os seguintes resultados:

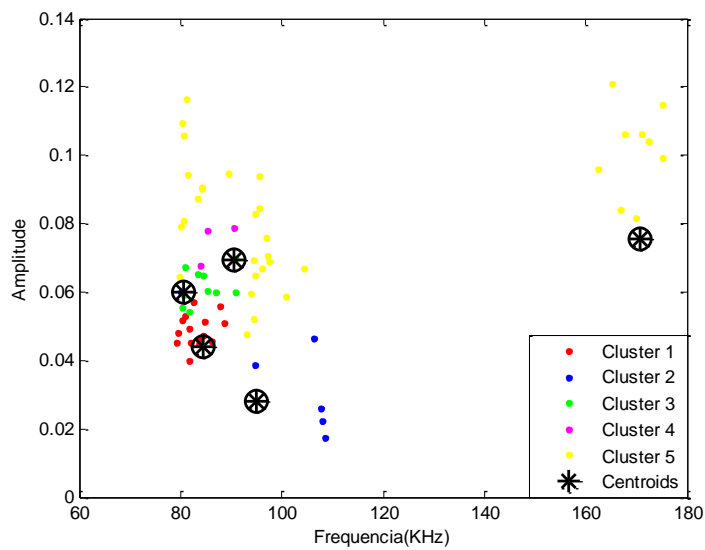


Figura 76. Classificação com os centroides do Caso 2.

Tabela 36. Índice de Acerto com os centroides do Caso 2.

Ítem	Classe	Ondas PE	Ondas PI
1	Cluster 1	10	5
2	Cluster 2	1	5
3	Cluster 3	10	0
4	Cluster 4	3	0
5	Cluster 5	10	24
Índice de Acerto		72,06%	

O caso 3 finalizou a classificação com 4 classes e com um índice de acerto de 88,89%. Onde os *clusters* 1 e 4 definem as ondas de PI; e os *clusters* 2 e 3 as ondas de PE.

Ao utilizar os centroides destas 4 classes e classificar o conjunto maior de dados obtiveram-se os seguintes resultados:

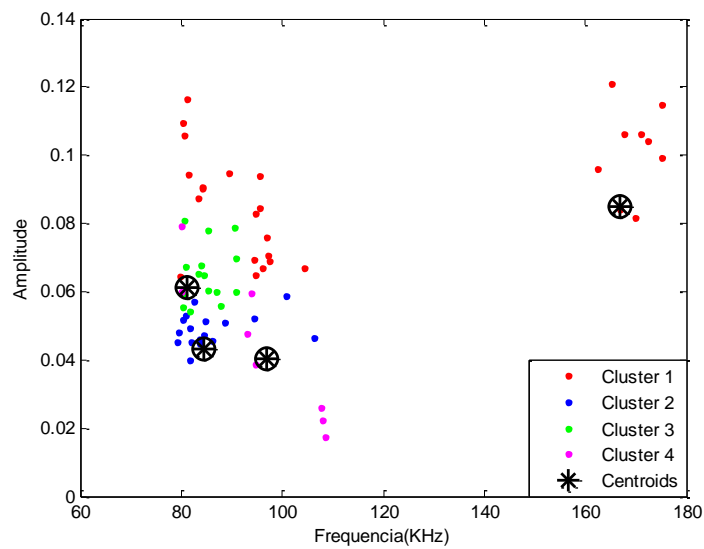


Figura 77. Classificação com os centroides do Caso 3.

Tabela 37. Índice de Acerto com os centroides do Caso 3

Item	Classe	Ondas PE	Ondas PI
1	Cluster 1	8	20
2	Cluster 2	12	5
3	Cluster 3	14	1
4	Cluster 4	0	8
Índice de Acerto		79,41%	

Por último, o caso 4 finalizou a classificação com 5 classes e com um índice de acerto de 86,11%. Onde os *clusters* 1 e 3 definem as ondas PE; e os *clusters* 2, 4 e 5 as ondas de PI.

Ao utilizar os centroides destas 5 classes e classificar o conjunto maior de dados, obtiveram-se os seguintes resultados:

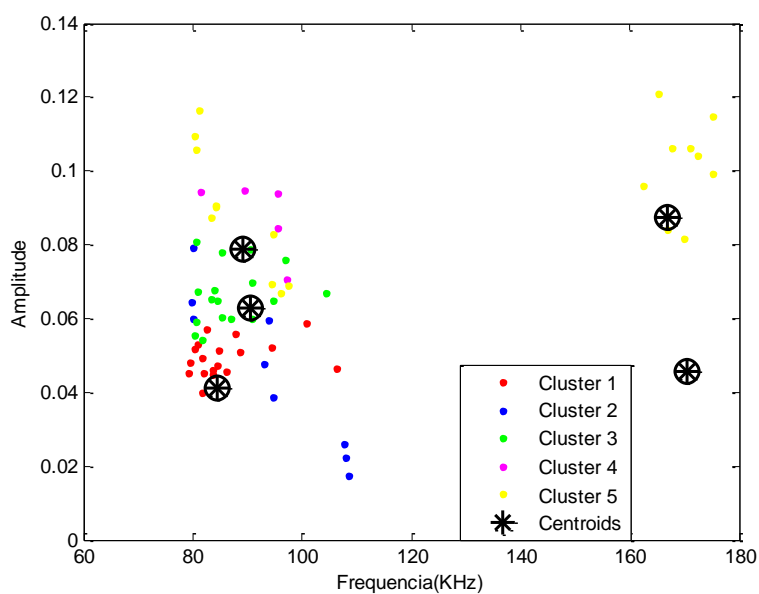


Figura 78. Classificação com os centroides do Caso 4.

Tabela 38. Índice de Acerto com os centroides do Caso 4.

Ítem	Classe	Ondas PE	Ondas PI
1	<i>Cluster 1</i>	12	6
2	<i>Cluster 2</i>	0	9
3	<i>Cluster 3</i>	16	1
4	<i>Cluster 4</i>	4	1
5	<i>Cluster 5</i>	2	17
Índice de Acerto		80,88%	

Os resultados obtidos com os centroides de cada caso da seção 4.2 são bastante satisfatórios, menos a do primeiro caso. A provável causa disso seja que, ao ter um alto índice de acerto, o classificador ficou muito específico para o caso 1.

4.6 Comparação de Resultados com outros trabalhos

Ao comparar os resultados obtidos no trabalho desenvolvido por Pinto em 2011 [13], para este mesmo ensaio ele considerou 3 casos. No primeiro a classificação foi feita para duas classes, sem propagação (SP) e com propagação (P), tendo uma taxa de acerto de 98%. No segundo caso a classificação tomou em conta três classes sem reclassificação, sem propagação (SP), com propagação estável (PE) e com propagação instável (PI). E no terceiro caso foram consideradas as mesmas classes com reclassificação. As taxas de acerto destes dois últimos foram de 86%, que são semelhantes às obtidas nesta pesquisa; mas com a diferença de que aqui a classificação foi feita em duas classes (PE e PI).

O provável motivo dos resultados serem semelhantes pode radicar na complexidade de separar os sinais de SP e de PE.

5 Considerações

Para a continuação desta pesquisa ou para o desenvolvimento de um trabalho semelhante é importante ter as seguintes considerações:

1. Na implementação de um sistema de aquisição, é importante ter um bom entendimento do processo para poder dimensionar a instrumentação e os sistemas de processamento de dados, para garantir a precisão do sistema.
2. Nos casos de sistemas com altas taxas de aquisição e de armazenamentos de dados é fundamental a partição desses dados, para não gerar megadados que depois não possa ser processada pela limitação de memória dos processadores.
3. É importante ter em conta durante o desenvolvimento de classificadores a medição da acurácia dos mesmos, para poder evitar o *overfitting* ou *overtraining* e manter a capacidade de generalização. E desta forma potencializar o custo computacional.

6 Conclusões

A partir do presente estudo e desenvolvimento de classificadores para detecção de trincas, a partir da forma de onda EA, chegou-se às seguintes conclusões:

1. Os classificadores desenvolvidos para o conjunto de dados menor mostraram-se ser bastante eficientes, e com alta capacidade de generalização. Esta eficiência está relacionada fortemente à boa separação das ondas de PE e de PI no momento da escolha dos dados.
2. Os classificadores desenvolvidos para o maior conjunto de dados mostraram-se muito específicos e pouco eficientes no momento da generalização. A causa disto, embora trabalhar com uma maior quantidade de dados ajudaria a generalizar melhor, se deve a que neste conjunto de dados foram incluídas as formas de onda que se encontram na fase de transição entre a propagação estável e a propagação instável da trinca.
3. Baseados nos resultados obtidos na seção 4.5, pode-se concluir que a classificação inicializando os centroides mostrou-se levemente mais eficiente que a inicialização aleatória dos centroides. E que a classificação com reclassificação é levemente mais eficiente que a classificação sem reclassificação.
4. Os resultados apresentados neste trabalho acusam que a detecção de trincas e identificação da fase da mesma usando sensores de EA é viável para ser implementado em ambientes industriais. E a simplicidade do processamento do *K-means* ajudaria na implementação de um sistema *on-line*.

7 Trabalhos Futuros

Como trabalhos futuros, pode-se ter as seguintes sugestões:

1. Melhoria do sistema de aquisição visando um sistema capaz de adquirir sinais com uma quantidade maior de sensores a uma taxa de amostragem de 1MSa/s. Assim como a capacidade de processamento e armazenamento.
2. Realizar ensaios em dutos de maior comprimento para poder identificar a distância apropriada à qual devem ser instalados os sensores para otimizar recursos.
3. Obter novos e maiores bancos de dados para tentar melhorar a acurácia de classificação e a generalização. Assim, como testar outros classificadores não supervisionados.
4. Desenvolvimento de um sistema supervisorio em tempo real para detecção de defeitos em dutos de transporte de petróleo ou gás. Para o qual é importante o dimensionamento da unidade de processamento, sendo indispensável uma alta velocidade para isso, o que provavelmente implique o uso de *clusters*.

8 Referência Bibliográfica

- [1] B. B. Matos, "Detecção de Falhas Induzidas pelo Hidrogênio via Emissão Acústica" *Tese* - Universidade Federal do Rio de Janeiro, COPPE, 2004.
- [2] S. D. Soares, "Correlações entre Emissão Acústica de Defeitos e Parâmetros de Mecânica de Fratura na Avaliação da Integridade de Equipamentos Pressurizados" *Tese* - Universidade Federal do Rio de Janeiro, COPPE, 2008.
- [3] C. F. C. Pinto, "Monitoração de Defeitos em Dutos Rígidos por Emissão Acústica e Redes Neurais" *Disertação* - Universidade Federal do Rio de Janeiro, COPPE, 2011.
- [4] Vallen Systeme, "Acoustic Emission Preamplifiers Specification", Released 08-2012.
- [5] Mistras, "R15I-AST Sensor Integral Preamplifier Sensor" 2011.
- [6] Mistras, "AE Sensors & Preamplifiers Users Manual Rev 0," June 18, 2002.
- [7] National Instruments, "NI 625x Specifications" June, 2007.
- [8] J. M. de Moraes, "Petróleo em águas Profundas, Uma história tecnológica da Petrobras na exploração e produção offshore" - IPEA - 2013
- [9] A. Sidorova, "Aplicación del Método de la Emisión Acústica en la Monitorización de las Estructuras de Hormigón", *Disertação* - Universitat Politècnica de Catalunya, 2010.
- [10] S. Haykin, "Neural Networks: A Comprehensive foundation" - 2nd Ed - Prentice Hall, 1999
- [11] R.O Duda, P. E Hart, D. G Stork, "Pattern Classification"- 2nd Ed - Wiley, 2001.
- [12] L. Nunes, "Análise e Síntese de Estratégias de Aprendizado para Redes Neurais Artificiais", *Disertação* - Universidade Estadual de Campinas.
- [13] V. J. Frainer, "Avaliação do Comportamento em Fratura de Riser Rígidos", *Tese* - Universidade Federal do Rio Grande do Sul, 2007.
- [14] G. A. Tarnowski, "Determinação da Pressão de Colapso em Tubos API 5L X56 aplicando o Método R6", *Dissertação* - Universidade Federal do Rio Grande do Sul, 2003.
- [15] C. F. C. Pinto, "Monitoração de Defeitos em Dutos Rígidos por Análise dos Parâmetros de Emissão Acústica Utilizando Redes Neurais", *Tese* - Universidade Federal do Rio de Janeiro, COPPE, 2014.

- [16] <http://ocean-noise.com/blog/2013/04/noise-awareness-day-gets-wet/>, acessado em maio de 2014.
- [17] <http://novos-cientistas.blogspot.com.br/2012/05/redes-neurais.html>, acessado em outubro de 2015.
- [18] Relatórios do Projeto "Monitoramento de Defeitos em Dutos Rígidos por Emissão Acústica e Redes Neurais", CENPES, 2012.
- [19] http://www.sas.com/pt_br/insights/big-data/what-is-big-data.html, acessado em novembro de 2015.

APÊNDICE

A. Megadados

Em tecnologia da informação, o termo Megadados refere-se a um grande conjunto de dados armazenados. O analista Doug Laney definiu, em 2001, os Megadados em três Vs: Volume, Velocidade e Variedade; mas atualmente também são consideradas a Variabilidade e a Complexidade. A seguir explicar-se-á cada um destes termos:

1. **Volume:** Muitos fatores contribuem para o aumento do volume de dados. Transações de dados armazenados ao longo dos anos, dados de texto constantemente em streaming nas mídias sociais, o aumento da quantidade de dados de sensores que estão sendo coletados etc. No passado o volume de dados excessivo criou um problema de armazenamento. Mas com os atuais custos de armazenamento decrescentes, outras questões surgem, incluindo, como determinar a relevância entre os grandes volumes de dados e como criar valor a partir dessa relevância.
2. **Velocidade:** De acordo com o Laney, velocidade significa tanto o quão rápido os dados estão sendo produzidos quanto o quão rápido os dados devem ser tratados para atender a demanda. Etiquetas RFID e contadores inteligentes estão impulsionando uma necessidade crescente de lidar com torrentes de dados em tempo quase real. Reagir rápido o suficiente para lidar com a velocidade é um desafio para a maioria das organizações.
3. **Variedade:** Os dados de hoje vêm em todos os tipos de formatos. Sejam bancos de dados tradicionais, hierarquias de dados criados por usuários finais e sistemas OLAP, arquivos de texto, e-mail, medidores e sensores de coleta de dados, vídeo, áudio, dados de ações do mercado e transações financeiras. Por algumas estimativas, 80 por cento dos dados de uma organização não é numérico! Mas, estes dados também precisam ser incluídos nas análises e tomadas de decisões das empresas.
4. **Variabilidade:** Para além das velocidades crescentes e variedades de dados, o fluxo de dados pode ser altamente inconsistente com picos periódicos. Há algo grande virando tendência nas mídias sociais? Talvez haja um grande IPO se aproximando. Talvez de repente nadar com animais exóticos nas Bahamas se torne a atividade de férias imperdível do momento. Diariamente, eventos sazonais desencadeiam picos de carga de dados e isso pode ser um desafio para gerenciar - especialmente quando temos as mídias sociais envolvidas.
5. **Complexidade:** Quando você lida com grandes volumes de dados, eles vêm de diversas fontes. É um grande desafio vincular, correlacionar, limpar e transformar os dados de um sistema. No entanto, é necessário conectar e correlacionar interações, hierarquias e vínculos múltiplos de informação ou então os dados podem rapidamente sair de controle. Governança de dados pode ajudar a determinar como os dados díspares se relacionam com definições comuns e como integrar sistematicamente os ativos de dados estruturados e não estruturados para produzir informações de alta qualidade, uteis, adequadas e atualizadas.