

**1º TERMO ADITIVO AO ACORDO DE
COOPERAÇÃO CIENTÍFICA E TECNOLÓGICA
FIRMADO ENTRE VALE S.A., A UFES E A FEST**

Pelo presente instrumento de um lado a **VALE S.A.**, sociedade sediada na Praia de Botafogo nº 186, Rio de Janeiro – RJ, CEP 22.250-145, inscrita no CNPJ/MF sob o nº 33.592.510/0001-54, adiante denominada **VALE**, aqui representada por seus representantes legais infra assinados, e, de outro lado, a **UNIVERSIDADE FEDERAL DO ESPÍRITO SANTO**, doravante denominada **UNIVERSIDADE**, autarquia educacional de regime especial, situada na Av. Fernando Ferrari, 514, Campus Universitário, Goiabeiras, Vitória, ES, CEP 29.075-910, inscrita no CNPJ/MF sob o nº. 32.479.123/0001-43, neste ato representada pelo seu Reitor, Prof. Paulo Sérgio de Paula Vargas, brasileiro, solteiro, portador da carteira de identidade nº. 337.068 SSP ES, CPF nº. 526.372.397-00, adiante denominada **UFES**, e com interveniência da **FUNDAÇÃO ESPÍRITO SANTENSE DE TECNOLOGIA**, doravante denominada **FEST**, Fundação de Direito Privado sem fins lucrativos, com sede na Av. Fernando Ferrari, nº 845 – Campus Universitário, Goiabeiras, Vitória – ES, inscrita no CNPJ/MF sob o nº 02.980.103/0001-90, representada neste ato pelo Superintendente, Armando Biondo Filho, brasileiro, divorciado, portador da carteira de identidade 3.052.172 IFP- RJ, CPF nº 376.717.407-30, adiante denominada **FUNDAÇÃO**, individualmente denominadas “Parte” e em conjunto “Partes”, CONSIDERANDO que, em 18/06/2020, as Partes celebraram o Acordo de Cooperação Científica e Tecnológica, doravante denominado “Acordo”, para execução do Projeto de Pesquisa: “**Desenvolvimento de uma Ferramenta de Software orientada à auditoria de comunicação via Rádio;**

CONSIDERANDO que as Partes mantêm a relação jurídica em condições de pleno equilíbrio;

CONSIDERANDO o interesse das Partes em prorrogar o prazo do Acordo e alterar o Anexo I.

Resolvem celebrar o presente 1º Termo Aditivo ao Acordo (“Termo Aditivo”), de acordo com as seguintes cláusulas e condições:

CLÁUSULA PRIMEIRA – DO OBJETO

1.1. O presente Termo Aditivo tem como objeto a prorrogação do prazo do Acordo por mais 05 (cinco) meses daquele constante da Cláusula Sétima, bem como a substituição do Anexo I do Acordo.

CLÁUSULA SEGUNDA – DAS ALTERAÇÕES

2.1. Em consequência do disposto na cláusula 1.1 acima, a Cláusula Sétima do Acordo passará a vigorar com a seguinte redação:

CLÁUSULA SETIMA: DA VIGÊNCIA

O presente ACORDO vigorará pelo prazo de 33 (trinta e três) meses, a partir da data de sua assinatura, extinguindo-se após o cumprimento de todas as suas obrigações, sendo certo que a cláusula de Propriedade Intelectual, terá vigência de 20 (vinte) anos e as de confidencialidade pelo prazo de 10 (dez) anos a contar do encerramento do ACORDO.

2.2 Em consequência do disposto na cláusula 1.1 acima, fica alterado o Anexo I do Acordo pelos anexos do presente aditivo, a saber: formulário de propostas e planilha de orçamento, ambos devidamente rubricados pelas Partes.

CLÁUSULA TERCEIRA – DAS DISPOSIÇÕES GERAIS

3.1. As Partes, através do presente Termo Aditivo, dão a mais plena, geral, rasa e irrevogável quitação, para todos os fins de direito, por todos os fatos passados até a presente data, ratificando todos os atos praticados e nada mais tendo a reivindicar, em juízo ou fora dele, a qualquer título, em relação às obrigações contratuais até aqui já executadas.

3.1.1 A quitação outorgada no item 3.1 acima não se aplica às garantias legais e/ou contratuais, bem como as demais responsabilidades das Partes que, por sua natureza tenham caráter perene ou prazo prescricional ainda não decorrido, especialmente as relativas à responsabilidade civil perante terceiros, encargos trabalhistas e previdenciários, obrigações fiscais, direitos de propriedade intelectual e obrigação de confidencialidade, bem como a qualquer pleito futuro baseado em fatos desconhecidos pela outra Parte na data do presente Termo Aditivo

3.2. Permanecem inalteradas e ratificadas todas as demais Cláusulas do Acordo, naquilo em que não conflitarem com o teor deste instrumento.

Em caso de assinatura física, o Contrato será assinado em 3 (três) vias de igual teor e forma, para um só efeito. Como alternativa à assinatura física do Contrato, as Partes declaram e concordam que a assinatura mencionada poderá ser efetuada em formato eletrônico. As Partes reconhecem a veracidade, autenticidade, integridade, validade e eficácia deste Contrato e seus termos, incluindo seus anexos, nos termos do art. 219 do Código Civil, em formato eletrônico e/ou assinado pelas Partes por meio de certificados eletrônicos, ainda que sejam certificados eletrônicos não emitidos pela ICP-Brasil, nos termos do art. 10, § 2º, da Medida Provisória nº 2.200-2, de 24 de agosto de 2001 (“MP nº 2.200-2”).

Espírito Santo, 07 de Junho de 2021

VALE S.A.

Nome: André Rezende Soares
CPF: 043.677.957-99

VALE S.A.

Nome: Rafael Gaier Gomes
CPF: 054.073.197-80

UFES

Nome: Paulo Sérgio de Paula Vargas

FEST

Nome: Armando Biondo Filho

Testemunhas:

Nome: Jorge Leonid Aching Samatelo
CPF: 059.930.517-79

Nome: Henrique Rieveres Borges de Andrade
CPF: 884.547.897-15

Este documento foi assinado eletronicamente por Paulo Sérgio de Paula Vargas, Jorge Leonid Aching Samatelo, André Rezende Soares, Rafael Gaier Gomes, Henrique Rieveres Borges de Andrade e Armando Biondo Filho.
Para verificar as assinaturas vá ao site <https://vale.portaldeassinaturas.com.br:443> e utilize o código 74DF-171E-AECO-78B5. This document has been electronically signed by Paulo Sérgio de Paula Vargas, Jorge Leonid Aching Samatelo, André Rezende Soares, Rafael Gaier Gomes, Henrique Rieveres Borges de Andrade and Armando Biondo Filho. To verify the signatures, go to the site <https://vale.portaldeassinaturas.com.br:443> and use the code 74DF-171E-AECO-78B5.

Desenvolvimento de uma Ferramenta de Software orientada à auditoria de comunicação via Rádio

Dr. Jorge Leonid Aching Samatelo

FORMULÁRIO DE APRESENTAÇÃO DE PROPOSTAS

[Vitoria, 12/05/2021]
[1.6]

Sumário

1. Identificação	4
1.1 Dados do Proponente	4
1.2 Área da Vale	4
2. Dados do Projeto	5
3. Equipe do Projeto	5
4. Palavras Chave do Projeto	6
5. Resumo do Projeto de Pesquisa	6
6. Justificativa.....	6
7. Descrição do Estado da Arte	8
7.1 Grau de maturidade da tecnologia a ser desenvolvida.....	9
8. Objetivos	10
8.1 Gerais.....	10
8.2 Específicos	10
9. Metodologia de Pesquisa.....	10
10. Resultados Esperados.....	11
11. Grau de inovação do projeto	12
11.1 Justificativa do grau de inovação	12
12. Possibilidade de patenteamento	12
13. Acesso à Vale.....	12
14. Riscos	12
15. Relevância estratégica para Vale.....	13
15.1 Crescimento de Mercado – Foco em vendas.....	13

Este documento foi assinado eletronicamente, por Paulo Sérgio de Paula Vargas; Jorge Leonid Aching Samavolo, André Resende Soares, Rafael Gaier Gomes, Henrique Rieveres Borges de Andrade e Armando Biondo Filho.
 Para verificar as assinaturas vá ao site <https://vale.portaldeassinaturas.com.br:443> e utilize o código 74DF-171E-AECC-78B5. This document has been electronically signed by Paulo Sérgio de Paula Vargas; Jorge Leonid Aching Samavolo, André Resende Soares, Rafael Gaier Gomes, Henrique Rieveres Borges de Andrade e Armando Biondo Filho. To verify the signatures, go to the site <https://vale.portaldeassinaturas.com.br:443> and use the Code 74DF-171E-AECC-78B5.

15.2	Redução de Custos – Foco em melhoria de processo	13
15.3	Implicações ambientais.....	14
15.4	Implicações em saúde e segurança.....	14
16.	Cronograma de Atividades e Marcos	14
17.	Produtos e Entregas.....	16
18.	Referências Bibliográficas da Pesquisa	17
19.	Orçamento Detalhado e Cronograma de Desembolso	19
20.	Informações Adicionais.....	19
21.	Anexos.....	19
22.	Assinaturas	19

Este documento foi assinado eletronicamente por Paulo Sérgio de Paula Vargas, Jorge Leonid Aching Samatelo, André Resende Soares, Rafael Gaier Gomes, Henrique Rievers Borges de Andrade e Biondo Filho.
 Para verificar as assinaturas vá ao site <https://vale.portaldeassinaturas.com.br:443> e utilize o código 74DF-171E-AECC0-78B5. This document has been electronically signed by Paulo Sérgio de Paula Vargas, Jorge Leonid Aching Samatelo, André Resende Soares, Rafael Gaier Gomes, Henrique Rievers Borges de Andrade and Biondo Filho. To verify the signatures, go to the site <https://vale.portaldeassinaturas.com.br:443> and use the code 74DF-171E-AECC0-78B5.

1. Identificação

1.1 Dados do Proponente

Instituição:	Universidade Federal de Espírito Santo - UFES
Nome do Pesquisador:	Jorge Leonid Aching Samatelo
CPF:	059.930.517-79
Nacionalidade:	Peruano
Titulação:	() Graduado () Especialista () Mestre () Doutor (x) Pós-Doutorado
Telefone:	(27) 981487705
Celular:	(27) 981487705
E-mail:	jorge.samatelo@ufes.br
Departamento/ Unidade:	Engenharia Elétrica
Área de Formação/ Especialização:	Engenharia Eletrônica
Endereço:	Avenida Alziro Zarur 310, Ap 209, Bairro Jardim da Penha
Cidade:	Vitoria
Estado:	Espírito Santo
CEP:	29060-350
País:	Brasil

Caso o proponente não seja o coordenador do projeto, informar seus dados:

1.2 Área da Vale

Área da Vale envolvida:	Engenharia EFVM
Contato:	Henrique Rieveres Borges de Andrade
Telefone:	(27) 3333-7496
E-mail:	Henrique.andrade@vale.com

2. Dados do Projeto

Título do Projeto:		Desenvolvimento de uma Ferramenta de Software orientada à auditoria de comunicação via Rádio	
Duração (em meses):		29	
Projeto em Rede:		(X) Individual () Rede*	
Programa/ Linha de Pesquisa**:		Processamento Digital de Sinais – PDS	
Tipo de Pesquisa:		() Pesquisa Básica (X) Pesquisa Aplicada () Desenvolvimento () Transferência de Tecnologia	
Aplicável a Lei do Bem:		(X) Sim () Não	
Versão	Data	Autor	Alteração
1.0	29/09	Jorge	Versão 1
1.1	23/10	Henrique	Versão 1_ Comentado Vale
1.2	21/11	Jorge	Versão 2
1.3	28/11	André	Versão 2_ Comentado Vale (Comentários como sugestão), alterações nas revisões do documento diretamente no texto.
1.4	12/12	Jorge	Versão 3
1.5	13/12	André	Versão 4 – Ajustes finais: Incluído equipe Vale.
1.6	12/05/2021	Jorge	Versão 5 – Modificação para adição de aditivo de 5 meses

*Projeto relacionado com um ou mais projetos.

**No âmbito das linhas de pesquisa apresentadas pela Vale.

3. Equipe do Projeto

Instituição	Nome	Titulação	Telefone	E-mail	Participação no Projeto e Função	Link no Currículo Lattes
UFES	Jorge Leonid Aching Samatelo	Doutorado	(27) 981487705	jorge.samatelo@ufes.br	Coordenador	http://lattes.cnpq.br/509258096050209
UFES	Raquel Frizera Vassallo	Doutorado	(27) 992946309	raquel@ele.ufes.br	Pesquisador	http://lattes.cnpq.br/572903915280374
UFES	Daniel Jesus Ribeiro	Graduação	(27) 981487705	daniel.ribeiro@vale.com	Aluno	http://lattes.cnpq.br/1252943175213470
UFES	Gabriel Soares Xavier	Graduação	(27) 996628655	gabriel.s.xavier@edu.ufes.br	Aluno	http://lattes.cnpq.br/7492668333512575
UFES	João Victor dos Santos Nunes	Graduação	(27) 992859200	joavictor.santosnunes@gmail.com	Aluno	http://lattes.cnpq.br/4905019096515194
UFES	Lucas Grigoletto Scart	Mestrado	(27) 981887669	scart.lucas@gmail.com	Aluno	http://lattes.cnpq.br/21742300685268
UFES	Eddy Giusepe Chirinos Isidro	Doutorado	(27) 997455087	eddychirinos.unac@gmail.com	Aluno	http://lattes.cnpq.br/49059201985220660
Vale	Henrique Rievers Borges de Andrade	Especialização	(27) 999571108	henrique.andrade@vale.com	Pesquisador	

Vale	André Stanzani Franca	Mestrado	(27) 999927906	andre.franca@vale.com	Pesquisador	http://lattes.cnpq.br/4403628127242366
------	-----------------------	----------	----------------	-----------------------	-------------	--

4. Palavras Chave do Projeto

Auditoria de comunicações via rádio; Processamento Automática de Fala; *Deep Learning*;

5. Resumo do Projeto de Pesquisa

CONTEXTUALIZAÇÃO. Parte do funcionamento da ferrovia da VALE depende do cumprimento das diretrizes e padrões estabelecidos para a comunicação via rádio entre os usuários da ferrovia. Especificamente entre os condutores do equipamento ferroviário e os controladores de tráfego central e local. Para garantir o cumprimento das normas é efetuado um processo de auditoria da comunicação por voz, que analisa a eficácia da comunicação estabelecida via rádio, bem como a inteligibilidade da fala do mesmo. Atualmente, tal procedimento é manual gerando uma inversão tanto de tempo como pessoal capacitado,

PROPOSTA. Neste sentido, aqui é proposto o desenvolvimento de uma ferramenta computacional de suporte a auditoria orientada ao atendimento dos inspetores do CCO/CCP (Centro de controle operacional/Centro de controle de pátio) que avaliam a comunicação estabelecida via. Tal ferramenta permitirá efetuar uma transcrição do áudio bem como diferenciar dois interlocutores dentro de uma mesma gravação.

METODOLOGIA. Para tal fim, serão desenvolvidos três modelos computacionais baseados em redes neurais profundas, que na atualidade definem o estado da arte para aplicações de processamento de fala. De maneira específica, o primeiro modelo estará orientado ao reconhecimento automático de fala em português. O segundo direcionado à segmentação de falante e identificação automática de oradores. Os dois modelos são de natureza supervisionada, implicando na necessidade de uso de bancos de dados públicos de áudios em português, além da necessidade de desenvolver um banco de áudio de interlocutores a ser elaborado no projeto.

RESULTADOS. Todos os modelos desenvolvidos serão integrados em uma aplicação de software orientada ao uso no processo de auditoria e capacitação. Para um correto funcionamento da aplicação seu desenvolvimento será feito em etapas, especificamente, integração dos modelos, desenvolvimento da interface de usuário, testes e melhoras.

6. Justificativa

A segurança da ferrovia é baseada em sistemas que permitem ou não o tráfego de uma composição em um determinado trecho da malha ferroviária denominado circuito. Os centros de controle são os responsáveis por reservar um circuito para o trânsito de uma composição, bem como gerenciar o tráfego de todos os trens na linha tronco e pátios de manobra. A liberação dos trens ocorre através de comunicação via rádio. Buscando um maior grau de confiabilidade, são definidos padrões e protocolos de comunicação que contemplam situações de passagem de informações, liberações de movimentação, atualização de restrição, passagem de serviço entre outras rotinas.

Buscando garantir o bom funcionamento da ferrovia, existe o processo de auditoria da comunicação por voz que analisa a eficácia da comunicação estabelecida via rádio, bem como o cumprimento das diretrizes e padrões estabelecidos.

Atualmente, o procedimento de auditoria é executado por inspetores do CCO/CCP que avaliam a comunicação estabelecida pelo controlador que está sendo auditado. Verifica-se o cumprimento das normas, bem como a inteligibilidade da fala do mesmo. São pontuadas as falhas de comunicação e é fornecido um **feedback** dos pontos que podem ser melhorados.

O projeto busca o desenvolvimento de uma ferramenta que seja capaz de realizar duas atividades fundamentais:

1. Efetuar uma transcrição do áudio.
2. Diferenciar dois interlocutores dentro de uma mesma gravação.

Este documento foi assinado eletronicamente por André Franca, Henrique Rieveres Borges e Armando Blondo Filho. Para verificar as assinaturas vá ao site https://valedocuments.com.br/4403628127242366. O documento foi assinado eletronicamente por André Franca, Henrique Rieveres Borges e Armando Blondo Filho. To confirm the signatures, go to the site https://valedocuments.com.br/4403628127242366.

3. Criação da ferramenta computacional orientada ao suporte na auditoria dos inspetores do CCO/COB que avaliam a comunicação estabelecida via rádio.

Uma vez que a ferramenta esteja apta a fornecer tais informações, o procedimento de auditoria será simplificado: o auditor receberia a conversa em formato de texto, sendo atribuída a cada interlocutor a sua fala.

A implementação das referidas capacidades, implica de forma geral no desenvolvimento de modelos computacionais focados no processamento da fala (*speech processing*). De forma específica, no que tange à primeira prestação, será necessária a implementação de um reconhecedor de fala (*speech recognition*) em relação à segunda prestação implica no desenvolvimento de um segmentador (*speaker diarization*) e um identificador de falante (*speaker identification*). A continuação, de maneira resumida é descrito cada um destes elementos constitutivos.

Um reconhecedor automático da fala (*ASR - Automatic Speech Recognition*) refere-se a um sistema que permite o reconhecimento e a conversão da linguagem falada em texto. Também é conhecido como um conversor de fala em texto (*STT - Speech To Text*). A indústria desenvolveu uma ampla gama de produtos comerciais, onde o ASR como interface do usuário se tornou cada vez mais útil e difundido. As aplicações centradas no consumidor exigem cada vez mais que o ASR seja robusto para toda a gama de ruídos do mundo real e outras condições de distorção acústica. No entanto, o reconhecimento confiável de palavras faladas em ambientes acústicos realistas ainda é um desafio.

Um segmentador automático de falante (*ASD - Automatic Speaker Diarization*) refere-se à sistema que tem a capacidade de identificar um falante, e não reconhecer o que ele estão dizendo. Basicamente, o áudio de entrada é particionado em segmentos homogêneos de acordo com a identidade do orador. Assim, um ASD pode simplificar a tarefa de ASR ou pode ser usado para autenticar ou verificar a identidade de um falante como parte de um processo de segurança. A implementação de um ASD encontra diferentes desafios, por exemplo, o número de oradores nos áudios comumente é desconhecido, não há conhecimento prévio sobre a identidade das participantes, os oradores podem falar ao mesmo tempo podendo existir diferentes condições de gravação em cada áudio.

Um identificador automático de orador (*ASI - Automatic Speaker Identification*) refere-se à identificação de uma pessoa a partir das características da voz. Basicamente o sistema responde à pergunta: quem está falando? Um ASI pode simplificar a tarefa de um ASR treinado em vozes específicas ou pode ser usado para autenticar ou verificar a identidade de um falante detectado por um ASD. Existem duas formas de operação de um ASI (i) como autenticador de identidade, aqui o orador afirma ter uma certa identidade e a voz é usada para verificar essa afirmação; (ii) como verificador de identidade, aqui o ASI determina a identidade de um falante desconhecido. Em certo sentido, a autenticação do orador é uma correspondência 1:1, em que a voz de um orador é correspondida a um registro específico, enquanto a verificação do orador é uma correspondência 1:N, em que a voz é comparada com vários registros de voz. Em aplicativos forenses, é comum primeiro executar um processo de verificação para criar uma lista de "melhores correspondências" e, em seguida, executar uma série de processos de autenticação para determinar uma correspondência conclusiva.

Na atualidade, o desenvolvimento de um ASR, ASD ou ASI, se beneficiaram dos avanços de um campo de Inteligência Artificial denominado aprendizado profundo (DL – *deep learning*). Estes avanços são evidenciados não apenas pelo aumento de trabalhos acadêmicos publicados no campo, mais importante, pela adoção mundial da indústria de uma variedade de modelos de DL na implantação de sistemas para o processamento de áudio e fala. De forma específica, DL surge como uma bio-inspiração das estruturas neurais do córtex visual dos mamíferos (HUBEL; WIESEL, 1968). Os modelos de DL têm a capacidade de aprender o valor de seus parâmetros de maneira supervisionada (classificação de amostras, por exemplo) e não supervisionada (análise de padrões, por exemplo). Tal possibilidade permite realizar um procedimento em que se aprende com os dados, o que se mostrou eficiente em diversas linhas de pesquisa, inclusive na área processamento de áudio e fala. Técnicas de DL permitiu enfrentar com êxito a implementação de ASR comerciais em fala inglesa. Contudo, o custo computacional exigido por técnicas de DL é elevado, em termos de processamento (mais especificamente, demandam processamento paralelo em placas gráficas de alto desempenho - Graphics Processing Units - GPU). Entre as diferentes arquiteturas de redes neurais, as *Convolutional Neural Networks* - CNN (tradução livre, redes neurais convolucionais) têm tido desempenhos de destaque em aplicações de visão computacional e processamento de áudio. Outras arquiteturas de destaque são as *Long-Short Term Memory* - LSTM (HOCHREITER e SCHMIDHUBER, 1997) e as Gated Recurrent Unit - GRU (CHO et al, 2014), que são baseadas em redes neurais recorrentes. Ambas as arquiteturas, e suas variantes têm se mostrado

bastante eficazes em tarefas relacionadas a séries temporais, quando comparadas a técnicas clássicas de predição.

7. Descrição do Estado da Arte

A seguinte descrição do estado do Arte está dividida em três apartados, cada um direcionado a explicar as técnicas desenvolvidas na literatura para a elaboração de um ASR, ASD e um ASI.

Trabalhos relacionados ao desenvolvimento de um ASR

O desenvolvimento de um ASR teve avanços significativos com o uso de rede neurais profundas, no entanto, estas foram aplicadas apenas em um componente de um ASR padrão, especificamente, no extator de características. Assim, as redes neurais foram treinadas para classificar segmentos individuais de dados acústicos e suas distribuições de saída eram alimentadas como probabilidades para um modelo temporal probabilístico clássico como um HMM (*Hidden Model Markov*). Seguindo as tendências recentes em sistemas baseados em DL, propostas de ASR *end-to-end* tem aumentado ao longo dos anos [GRAVES 2012; MAIO e et al., 2015; MAAS e et al., 2015; AMODEI e et al., 2016]. Um ASR *end-to-end* é um sistema onde a maior parte de componentes de um ASR padrão são substituídas por uma única arquitetura de rede neural profunda. Assim, os parâmetros e características aprendidas por um único modelo são ajustados apenas pelo algoritmo de treinamento da rede aumentando a taxa de precisão do ASR. Os principais trabalhos que seguem esta linha são descritos a seguir.

Em [GRAVES e et al., 2006; GRAVES, 2012], é proposta a *Connectionist Temporal Classification* (CTC). O principal objetivo desse método é que o CTC é projetado especificamente para tarefas de classificação temporal, isto é, para problemas de rotulagem de sequências em que o alinhamento entre as entradas e os alvos é desconhecido. Também não requer dados de treinamento pré-segmentados ou pós-processamento externo para extrair a sequência de rótulos das saídas da rede. Desde então, o método CTC tem sido amplamente utilizado em sistemas de reconhecimento de fala de *end-to-end* e foi adotado pelo Google como o algoritmo padrão na Pesquisa por voz do Google.

Mais recentemente, redes neurais recorrentes foram aplicadas com sucesso ao reconhecimento de fala [CHOROWSKI e et al., 2015; BAHDANAU e et al., 2016]. Esses RNNs são baseados na arquitetura codificador-decodificador, geralmente usados para lidar com sequências de entrada e saída de comprimento variável necessária em tradução automatizada [BAHDANAU, 2014], geração de legendas de imagem [XU e et al., 2016] e síntese de escrita [GRAVES, 2013].

Trabalhos relacionados ao desenvolvimento de um ASD

Nos últimos anos, os ASD foram utilizados com sucesso para analisar a fala humana em vários cenários de dia a dia, desde telefonemas [SNYDER e et al., 2016] até reuniões de negócios [YELLA e et al., 2015] e transmissão de notícias [XU e et al., 2016]. Embora vários avanços no desenvolvimento de algoritmos de diarização de orador de alta qualidade tenham sido feitos nos últimos anos [LE LAN et al., 2016; WOUBIE e et al., 2016; XU e et al., 2016; FERRAS e et al., 2016], ainda há vários desafios a serem enfrentados, por exemplo, análise do discurso sobreposto ou modulações de voz do orador.

Um típico ASD geralmente consiste em quatro componentes: (i) Segmentação de fala, aqui o áudio de entrada é segmentado em seções curtas que se supõe ter um único orador e as seções que não são de fala são tratadas (ii) extração de *embeddings* de áudio, aqui características específicas como MFCCs [KENNY e et al., 2016] *speaker factors* [CASTALDO e et al., 2008] ou vetores-i [SHUM e et al., 2013; SENOUSSAOUI e et al., 2014; SELL e et al., 2014] são extraídos das seções segmentadas; (iii) Agrupamento, aqui o número de oradores é determinado, e as combinações de áudio extraídas são agrupadas nesses oradores; e opcionalmente (iv) ressegmentação [SELL e et al., 2015], onde os resultados do agrupamento são refinados para produzir os resultados finais da diarização. Atualmente, os *embeddings* de áudio baseados em redes neurais (vetores-d) têm sido amplamente utilizados em sistemas ASD, geralmente superando significativamente as técnicas baseadas em vetores-i. Cabe indicar que, a maioria desses ASD estão orientados à verificação de orador dependente de texto, onde as combinações de oradores são extraídas de palavras chaves específicas detectadas. Os principais trabalhos que seguem esta linha são descritos a seguir.

Em [ROUVIER e et al., 2015] é proposto o uso do conceito de *embedding* denominado como *speaker embeddings*, os quais são implementados via uma rede neural orientada a um problema de classificação, na qual a rede visa prever o identificador relacionada à identidade de cada orador. Essa abordagem atinge uma taxa de erro de diarização de 19,25% em amostras de áudio sobre um banco de dados de falantes rotulado

Em [Wan e et al., 2017] é proposto um *speaker embeddings*, baseado em redes LSTM, o modelo é treinado em segmentos de comprimento fixo extraídos de um grande corpus de discursos arbitrário. Eles mostram que os agrupamentos dos *embeddings* produzidos por essas redes geralmente superam significativamente as outras representações.

Trabalhos relacionados ao desenvolvimento de um ASI

Um ASI funciona a partir do princípio de que a voz de cada pessoa é única, como uma impressão digital, portanto, pode ser usada para identificar o orador ou autenticar sua identidade. Esses sistemas em geral analisam as características do discurso que são diferentes entre os oradores e são usadas em aplicativos para autenticação, vigilância e análise forense dos falantes. Assim, o objetivo de qualquer técnica usada em um ASI é extrair e utilizar características de boa qualidade que representam atributos intrínsecos do aparelho vocal de um orador. Existem várias abordagens apresentadas até agora para o reconhecimento de oradores, que incluem modelos generativos como GMM (*Gaussian Mixture Model*) [REYNOLDS e et al., 2000], modelos discriminativos baseados em kernels [CAMPBELL e et al., 2006]. Todas essas abordagens contribuíram para obter melhores taxas de reconhecimento de oradores. No entanto, o problema de extrair características invariantes específicas do orador robustas ao ruído ainda é um tópico aberto de pesquisa, pois a mistura de informações específicas do orador com outros componentes de informação da fala faz com que os ASI comprometam seu desempenho. Além disso, a influência da semelhança entre oradores, variabilidade de canal e ruído aditivo complica mais a situação [TURNER e et al., 2009]. Portanto, a necessidade de obter características que são insensíveis a essas corrupções é inevitável. Em alguns estudos como [LEE e OH, 2001; MALAYATH e et al., 2000], são feitos esforços para extrair informações específicas de orador a partir do sinal de fala, com a suposição de que a mensagem ou a informação linguística são componentes dominantes da informação e podem ser facilmente isoladas.

Os abordagens baseadas em DL para a tarefa de verificação de orador podem ser agrupadas em dois tipos: (i) baseados em extratores de características que fazem uso de características extraídas por uma CNN e os MFCC vetores-i ou vetores-d; (ii) propostas *end-to-end*. Essas duas abordagens diferem de várias maneiras. Na fase de treinamento, as características extraídas por uma CNN, denominadas como *Bottleneck Features*, são especificadas de maneira supervisionada de forma separada da etapa de reconhecimento. Enquanto em um abordagem *end-to-end*, tanto a etapa de extração como reconhecimento são treinados conjuntamente. O objetivo do treinamento das primeiras abordagens é aprender as características que discriminam os oradores. Mas em abordagens *end-to-end*, o sistema é treinado em pares para descobrir se as entradas correspondem ao mesmo orador. A qualidade e a quantidade dos pares de entrada são importantes para um sistema *end-to-end*. De acordo com [SNYDER e et al., 2016], um sistema de *end-to-end* treinado em um pequeno *dataset* não apresenta um bom desempenho, mas usando um *dataset* maior, os referidos sistemas superam as técnicas baseadas em extratores de características. Este resultado deve-se ao fato que sistemas *end-to-end* usam como entradas os segmentos de fala capturando melhor as estatísticas de cada falante. Por outro lado, os resultados relatados em [SNYDER e et al., 2017] mostram que os sistemas *end-to-end* apresenta desempenho pior quando comparado aos sistemas baseados em vetores-d e sistemas baseados em vetores-i quando são usados segmentos de áudio curtos para o treinamento.

7.1 Grau de maturidade da tecnologia a ser desenvolvida

- () **Tecnologia emergente:** o projeto visa o desenvolvimento de novas tecnologias que nunca foram aplicadas industrialmente (nova plataforma tecnológica ou inovação radical).
- (X) **Primeira aplicação na indústria, mas nenhuma solução dominante:** o projeto visa o desenvolvimento de tecnologias que já tenham sido aplicadas industrialmente de forma experimental por competidores da Vale, mas que ainda não chegaram ao nível de solução dominante na indústria mineral.
- () **Solução dominante, aberta a melhorias:** o projeto visa o desenvolvimento de melhorias incrementais em tecnologias que já atingiram o estágio de solução dominante na indústria mineral.

() **Tecnologia altamente explorada e difundida:** o projeto visa apoiar o processo de aplicação de tecnologias que são novas apenas para a Vale e que apresentam baixo potencial para melhorias incrementais.

() **Não se aplica**

8. Objetivos

8.1 Gerais

Este projeto tem dois objetivos principais:

1. Desenvolver e implementar modelos para: o reconhecimento de fala em português e para a segmentação e identificação de orador. Os modelos serão baseados em arquiteturas de rede neurais profundas consolidadas da literatura, os quais, inicialmente, serão treinados e testados com bancos de dados públicos, para logo, ser massivamente treinados e avaliados no banco de dados a ser desenvolvido, tal banco estará compostos por áudios de comunicação via rádio da operação da Vale e as respectivas transcrições.
2. Integrar os modelos desenvolvidos em uma ferramenta computacional de uso no processo de auditoria e capacitação de comunicações via rádio e telefone das gravações da operação da ferrovia da VALE.

8.2 Específicos

Os objetivos específicos do projeto são:

1. Uso de modelos baseados em redes CNN e recorrentes, orientadas ao desenvolvimento de um ASR, ASD e ASI. Os sistemas automáticos propostos devem ser *end-to-end*, e especializados para a fala em portuguesa.
2. Elaboração de um banco de dados de arquivos de áudio de interlocutores. Tal conjunto de dados deve estar composto de: áudios de comunicações via rádio e telefone gravados durante a operação da ferrovia e dever conter a informação a ser aprendida pelos modelos (início e fim de uma conversação, palavras faladas, interlocutores participantes, entre outros)
3. Construção da ferramenta computacional. Tal aplicação deve integrar os modelos desenvolvidos para os problemas em estudo, e ser orientada ao uso no processo de auditoria de comunicações via rádio e telefone das gravações da operação da ferrovia da VALE, além de incluir algumas funcionalidades e métricas de avaliação tais como: (i) palavras por segundo, intervalo de tempo adequado entre palavras e outras métricas; (ii) identificação da utilização da palavra “câmbio” ao final das frases e da identificação do veículo ferroviário e/ou do interlocutor; (iii) opções de filtro para seleção de interlocutor, data e período de tempo para se realizar a auditoria; (iv) entre outros a serem definidos no decorrer do projeto.
4. Formação de Recursos Humanos. Pretende-se produzir pelo menos duas dissertações de mestrado e dois trabalhos de conclusão de curso.

9. Metodologia de Pesquisa

A pesquisa a ser realizada no presente projeto enquadra-se no campo de pesquisa Teórico-Aplicada Quantitativa. Em sentido amplo de pesquisa, a investigação científica tratada neste trabalho tem a seguinte metodologia: (i) quanto à natureza, pesquisa aplicada; (ii) quanto aos objetivos, exploratórios e descritivos; (iii) quanto às abordagens, quantitativos; (iv) quanto aos procedimentos, experimentais. Baseado nestas questões o trabalho será desenvolvido conforme os passos apresentados a seguir. Esses passos englobam as atividades e ações principais a serem feitas.

O passo inicial é realizar um estudo teórico, o qual estará dividido em duas etapas. Na primeira etapa será efetuado um estudo de redes neurais profundas, com foco nas arquiteturas CNN e redes recorrentes. Na segunda etapa, cada aluno de mestrado se dedicará ao estudo dos métodos de processamento de fala, os modelos baseados em redes neurais profundas orientadas ao desenvolvimento de um ASR, ASD e ASI.

Embora seja feita uma dedicação maior no início da pesquisa para realizar esta tarefa, o estudo teórico das redes neurais profundas e das técnicas de processamento de fala continuará ao longo de quase todo o tempo do projeto na busca de se manter atualizado das principais novidades a respeito dos temas abordados.

Em paralelo, as especificações do projeto serão realizadas, iniciando com: a coleta dos dados e a definição dos protocolos de avaliação para cada problema em estudo (reconhecimento da fala, segmentação de falante e identificação de orador). A coleta de dados estará constituída de duas etapas explicadas a seguir:

(i) Procura de banco de dados públicos para cada problema em estudo. É de interesse contar com diferentes bases de dados, já que as redes neurais profundas têm um melhor desempenho na medida em que são treinadas com um maior conjunto de treinamento. (ii) Construção de um banco de dados referente à aplicação objetivo do projeto. Tal banco estará conformado por arquivos de áudio de comunicações via rádio e telefone gravados durante a operação da ferrovia. Cada áudio estará atrelado a um arquivo de metadados (xml) que contém a informação rotulada de interesse, como por exemplo: os tempos de início e fim de uma conversação, os tempos de início e fim de cada intervenção na comunicação de um interlocutor como um respectivo identificador de orador, as palavras faladas em cada instante de tempo, identificadores dos interlocutores participantes, identificadores de palavras chaves em uma comunicação via rádio, e outros tipos de informação que serão definidas no transcurso do desenvolvimento do projeto. Cabe indicar, que a correta rotulação dos dados, permitirá especificar os modelos na tarefa objetivo como também obter valores de desempenho admissíveis para uma aplicação a ser testada em campo.

Finalmente, contando já com os bancos de dados públicos e os protocolos de avaliação definidos, será iniciada a implementação dos modelos e terminando a elaboração do banco de dado referente à aplicação, os modelos já elaborados e testados serão adequados aos requerimentos do projeto, por último, será implementada a aplicação de software de suporte a auditoria.

Para realizar todas estas etapas, o grupo de pesquisa será dividido em dois equipes: a primeira equipe conformada pelos alunos de Iniciação Científica (IC) e a segunda equipe conformado pelos alunos de mestrado.

A primeira equipe será responsável da elaboração do banco de dados referente à aplicação objetivo do projeto. Especificamente: (i) deveram definir as ferramentas de *software* a usar para tal labor; (ii) determinar a estrutura do arquivo de metadados; (iii) fazer a tarefa de rotulação dos áudios; (iii) elaborar a ferramenta computacional. Também, serviram de apoio para outras tarefas a ser efetuadas pela segunda equipe.

Os trabalhos referentes à segunda equipe estão relacionados com: (i) a implementação de técnicas de pré-processamento de dados de áudio; (ii) o desenvolvimento dos modelos baseados em redes neurais para cada um dos problemas em estudo; (iii) implementação da API (*Application Programming Interface*) de programação.

A metodologia de pesquisa considerada para o desenvolvimento dos modelos é: (a) inicialmente serão implementados os modelos de maior relevância presentes na literatura, para depois serem propostas e implementadas novas modelos se for o caso; (b) cada um dos modelos será validado, usando para tal fim os bancos de dados selecionados e o protocolo de avaliação já definido; (c) os resultados obtidos serão analisados para obter um melhor entendimento do comportamento de cada modelo, com o intuito de efetuar uma etapa de ajustes e alterações e com isso repetição dos experimentos; (d) os modelos propostos serão comparadas com modelos de estado da arte, para se identificar os pontos fortes e fracos de cada método, e assim propor melhorias; (e) usando o banco de dados desenvolvido os modelos serão retreinados e testados na aplicação objetivo.

A metodologia para a implementação da ferramenta para o suporte à auditoria é: (a) os modelos desenvolvidos serão construídos usando a linguagem de programação *Python* e o *framework tensorflow* especializado para a elaboração de modelos de DL; (b) usando as prestações de integração de modelos de *tensorflow* e as extensões ferramentas para o desenvolvimento de *Python* uma API de programação será elaborada; (c) usando a API de programação uma interfase de usuário será desenvolvida.

10. Resultados Esperados

São esperados os seguintes resultados principais:

- Desenvolvimento de uma ferramenta computacional de suporte a auditoria orientada ao atendimento das necessidades dos auditores. Tal ferramenta permitirá: (i) uma diminuição do tempo dispêndio nos

O desenvolvimento da ferramenta computacional de suporte a auditoria permitirá, de forma simultânea, diminuir o tempo dispendido nos procedimentos de auditoria como o pessoal alocado no referido processo. Que, em geral, produzira uma maior eficiência do tempo empregado pelos controladores, resultando em uma comunicação mais eficiente entre os operadores das vias ferroviárias da VALE.

- Alta redução
- Moderada redução
- Pequena redução**
- Nenhuma redução

15.3 Implicações ambientais

Qual é o potencial de impacto de seu projeto nas condições ambientais ou redução do impacto ambiental causado por uma ou mais operações realizadas pela Indústria da Mineração ou por outra empresa de sua cadeia produtiva? Justifique

- Alto impacto positivo
- Moderado impacto positivo
- Impacto neutro**
- Impacto negativo

No caso de impacto positivo, assinale os tipos de implicações ambientais potenciais do projeto:

- Eficiência Energética
- Tratamento de resíduos
- Reuso de água
- Redução de emissões
- Preservação e recuperação
- Outra implicação. Qual?

15.4 Implicações em saúde e segurança

Qual é o potencial de impacto de seu projeto na redução dos riscos à integridade física e à saúde de trabalhadores envolvidos nas operações realizadas pela Indústria da Mineração, por outra empresa de sua cadeia produtiva ou pela comunidade do entorno? Justifique

- Alto impacto positivo
- Moderado impacto positivo**
- Impacto neutro
- Impacto negativo

No caso de impacto positivo, assinale os tipos de implicações em saúde e segurança potenciais do projeto:

- Segurança no trabalho**
- Saúde do trabalhador
- Doenças em geral
- Outra implicação. Qual?

16. Cronograma de Atividades e Marcos

#	Atividade	Início (mês)	Término (mês)
1	Contratação dos Alunos	1	5
2	Estudo Teórico	6	9
3	Especificações do Projeto	6	9
4	Treinamento em Comunicação para Equipe Projeto (Resp Vale)	8	8

5	Construção do banco de dados de áudio rotulado	9	15
6	Relatório parcial técnico e financeiro (Prestação de contas)	15	15
7.1	Desenvolvimento do modelo de Transcrição para Texto	10	19
7.2	Desenvolvimento do modelo para a Segmentação de Oradores	10	19
7.3	Desenvolvimento do modelo para a Identificação do Oradores	10	19
8	Apresentação dos resultados parciais	18	19
9	Teste e melhora dos modelos	15	24
10	Desenvolvimento da ferramenta computacional e Interface com usuários	22	29
11	Documentação	12	29
12	Relatório final técnico e financeiro (Prestação de contas)	29	29

Atv.	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	
1																														
2																														
3																														
4																														
5																														
6																														
7.1																														
7.2																														
7.3																														
8																														
9																														
10																														
11																														
12																														

Baseado nos objetivos específicos, o projeto será desenvolvido conforme as atividades apresentadas a seguir.

Atividade 1. Contratação dos Alunos: Tempo estipulado para efetivar a contratação dos alunos que compõem a equipe do projeto.

Atividade 2. Estudo teórico: (i) estudo sobre redes neurais profundas, principalmente as redes CNN, redes recorrentes e suas variantes; (ii) estudo de técnicas de pré-processamento de sinais de áudio de maior relevância na literatura; (iii) estudo de ferramentas para o rotulagem e transcrição de arquivos de áudio de interlocutores; (iv) estabelecer os requisitos do banco de dados a elaborar.

Atividade 3. Especificações do projeto: coleta de banco de dados públicos e definição do protocolo de avaliação (métricas, formas de avaliação, etc.) para o problema de (i) reconhecimento automático de fala; (ii) segmentação automática de falante; (iii) e identificação automática de orador.

Atividade 4. Treinamento em Comunicação na VALE: Capacitação dos membros do projeto nos protocolos de comunicação via rádio na operação de ferrovia, assim como dos processos de auditoria.

Atividade 5. Construção do banco de dados de áudio rotulado: Construção do banco de dados de áudio de comunicações via rádio e telefone gravados durante a operação da ferrovia, esta atividade implicará (i) obtenção dos áudios; (ii) elaboração/seleção de uma ferramenta especializada para a rotulação de áudio; (iii) rotulação dos áudios indicando a informação de interesse para os três problemas em estudo (tempos de início e fim de uma conversão, os tempos de início e fim de cada intervenção na comunicação de um interlocutor como um respectivo identificador de orador, as palavras faladas em cada instante de tempo, identificadores dos interlocutores participantes, identificadores de palavras chaves em uma comunicação via rádio, e outros tipos de informação que serão definidas no transcurso do desenvolvimento do projeto).

Atividade 6. Relatório parcial técnico e financeiro (Prestação de contas): entrega dos (i) relatório parcial com as atividades realizadas, os resultados alcançados e os produtos a serem entregues; (ii) relatório financeiro com os gastos do projeto relacionados e devidamente comprovados.

Atividade 7. Desenvolvimento de modelos baseados em redes neurais: para cada problema em estudo será efetuado: (i) a implementação dos algoritmos para o pré-processamentos dos áudios referentes aos bancos de dados selecionados; (ii) o desenvolvimento e implementação dos modelos baseados em redes CNN e redes recorrentes para cada problema.

Atividade 8. Apresentação de resultados parciais. Os resultados dos modelos desenvolvidos para cada problema em estudo serão apresentados, tomando em conta: (i) os testes em banco de dados públicos; (ii) os testes iniciais no banco de dados de áudios em desenvolvimento, (iii) testes com dados novos obtidos em campo.

Atividade 9. Teste e melhora dos modelos: para cada problema em estudo será efetuado: (i) a validação dos modelos propostos, usando: (a) o protocolo de avaliação definido e os bancos de dados públicos selecionados; (b) o banco de dados construído; (ii) ajustes e alterações dos modelos propostos e com isso repetição de experimentos; (iii) comparação dos modelos propostos com técnicas de estado da arte.

Atividade 10. Desenvolvimento da ferramenta computacional e interface com usuário: (i) integração dos modelos implementados para cada problema em uma API de programação; (ii) elaboração de uma aplicação que sirva de interfase entre a API desenvolvida e o usuário final. (iii) teste da ferramenta usando o banco de dados desenvolvido e testes feitos por usuários.

Atividade 11. Documentação: (i) preparação de artigos para conferências nacionais/internacionais e para periódicos relevantes com Qualis A ou B (ii) Documentação de uso das ferramentas e softwares desenvolvidos.

Atividade 12. Relatório final técnico e financeiro (Prestação de contas): entrega dos (i) relatório final com as atividades realizadas, os resultados alcançados e os produtos a serem entregues; (ii) relatório financeiro com os gastos do projeto relacionados e devidamente comprovados.

17. Produtos e Entregas

Considerando como produto a entregar o software a desenvolver, o seguinte cronograma

#	Produto	Descrição	Mês de Entrega	Responsável
1	Relatório parcial técnico e financeiro (Prestação de contas)	Relatório parcial com as atividades realizadas, os resultados alcançados e os produtos a serem entregues e o relatório financeiro com os gastos do projeto relacionados e devidamente comprovados	15	Jorge Leonid Aching Samatele
2	Banco de Dados a ser desenvolvido	O banco de dados a entregar estará conformado por arquivos de áudio de comunicações via rádio e telefone gravados durante a operação da ferrovia. Cada áudio estará atrelado a um arquivo de metadados (xml) que contém a informação rotulada de interesse, para o projeto.	16	Jorge Leonid Aching Samatele
3	API de programação	A API de programação a entregar incluirá: (i) a implementação das técnicas de pré-processamento de dados de áudio; (ii) os modelos baseados em redes neurais para cada um dos problemas em estudo; (iii) funções de suporte para integração dos modelos.	22	Jorge Leonid Aching Samatele

4	Ferramenta computacional 0.1	O primeiro protótipo da ferramenta computacional compreenderá: (i) uma interface de usuário simples; (ii) interfaces para carregar novos áudios; (iii) interface para a transcrição de áudios; (iv) interface para geração de relatórios personalizados.	26	Jorge Leonardo Aching Samatelo
5	Ferramenta computacional 1.0	O protótipo final da ferramenta computacional compreenderá: (i) a interface de usuário final; (ii) interfaces para carregar novos áudios; (iii) interface para a transcrição de áudios e reconhecimento de locutor.	29	Jorge Leonardo Aching Samatelo
6	Relatório final técnico e financeiro (Prestação de contas)	Relatório final com as atividades realizadas, os resultados alcançados e os produtos a serem entregues e o relatório financeiro com os gastos do projeto relacionados e devidamente comprovados	29	Jorge Leonardo Aching Samatelo

18. Referências Bibliográficas da Pesquisa

- AMODEI, D., ANUBHAI, R., BATTENBERG, E., et al. *Deep speech 2: end-to-end speech recognition in english and mandarin*. In: International Conference on Machine Learning, v. 48, pp. 1–10, New York, USA, June 2016.
- BAHDANAU, D., CHOROWSKI, J., SERDYUK, D., et al. *End-to-end attention-based large vocabulary speech recognition*. In: IEEE International Conference on Acoustics, Speech and Signal Processing, pp. 4945–4949, Shanghai, China, March 2016.
- BAHDANAU, D., CHO, K., BENGIO, Y. *Neural machine translation by jointly learning to align and translate*. In: International Conference on Learning Representations, pp. 1–15, San Diego, USA, April 2014.
- CAMPBELL, W., CAMPBELL, J., REYNOLDS, D., SINGER, E., TORRES-CARRASQUILLO, P., *Support vector machines for speaker and language recognition*. *Comput. Speech Lang.*, vol. 20, pp. 210-299, 2006.
- CASTALDO, F., COLIBRO, D., DALMASSO, E., LAFACE, P., VAIR, C., *Stream-based speaker segmentation using speaker factors and eigenvoices*. In *Acoustics Speech and Signal Processing, 2008. ICASSP 2008. IEEE International Conference on*. IEEE, 2008, pp. 4133–4136.
- CHOROWSKI, J., BAH DANAU, D., SERDYUK, D., et al. *Attention-based models for speech recognition*. In *Advances in Neural Information Processing Systems*, pp. 577–585, Montreal, Canada, December 2015.
- FERRAS, M., MADIKERI, S., MOTLICEK, P., BOURLARD, H., *System fusion and speaker linking for longitudinal diarization of tv shows*. In *ICASSP, 2016*.
- GRAVES, A., FERNÁNDEZ, S., GOMEZ, F. J., et al. *Connectionist temporal classification: Labeling unsegmented sequence data with recurrent neural networks*. In: International Conference on Machine Learning, pp. 369–376, Pittsburgh, USA, June 2006.
- GRAVES, A. *Supervised Sequence Labelling with Recurrent Neural Networks*. Studies in Computational Intelligence. Heidelberg, Germany, Springer Verlag, 2012.
- GRAVES, A. *Generating sequences with recurrent neural networks*. June 2013. Eprint arXiv:1308.0850v5
- HOCHREITER, S.; SCHMIDHUBER, J. *Long short-term memory*. *Neural Computation*. 9 (8): 1735–1780, 1997
- HUBEL, D. H.; WIESEL, T. N. *Receptive fields and functional architecture of monkey striate cortex*. *The Journal of Physiology*, v. 195, n. 1, p. 215–243, mar 1968. ISSN 0022-3751.
- KENNY, P., REYNOLDS, D., CASTALDO, F., *Diarization of telephone conversations using factor analysis*. In *IEEE Journal of Selected Topics in Signal Processing*, vol. 4, no. 6, pp. 1059–1070, 2010.

XU, Y., MCMLOUGHLIN, I., SONG, Y., WU, K. *Improved i-vector representation for speaker diarization*. *Circuits, Systems, and Signal Processing*, vol. 35, no. 9, pp. 3393–3404, 2016.

YELLA. S. H., *Speaker diarization of spontaneous meeting room conversations*. Ph.D. dissertation, Ecole Polytechnique Fédérale de Lausanne, 2015

19. Orçamento Detalhado e Cronograma de Desembolso

Preencha o formulário em excel com detalhamento do orçamento e cronograma de desembolso do projeto.

20. Informações Adicionais

Principais equipamentos já existentes necessários ao projeto (máximo de 5)*

Item	Quantidade	Local

*Este campo será utilizado para fins de registro.

Auxílio recebido ou solicitado a outras entidades para o projeto (indicar moeda)*

Entidade	Valor solicitado	Valor aprovado

*Bolsas de pesquisa, recursos financiados por agências de fomento, entre outros.

21. Anexos

#	Anexo	Descrição
1	Formulário de Orçamento	Formulário detalhado do orçamento da proposta de projeto de P&D
2		
3		
4		
5		

22. Assinaturas

Preparado por:

Jorge Leonid Aching Samatelo

Aprovado por:

Henrique Rieveres Borges de Andrade

PROTOCOLO DE ASSINATURA(S)

O documento acima foi proposto para assinatura digital na plataforma Portal de Assinaturas Vale. Para verificar as assinaturas clique no link: <https://vale.portaldeassinaturas.com.br/Verificar/74DF-171E-AEC0-78B5> ou vá até o site <https://vale.portaldeassinaturas.com.br:443> e utilize o código abaixo para verificar se este documento é válido. The above document was proposed for digital signature on the platform Portal de Assinaturas Vale . To check the signatures click on the link: <https://vale.portaldeassinaturas.com.br/Verificar/74DF-171E-AEC0-78B5> or go to the Website <https://vale.portaldeassinaturas.com.br:443> and use the code below to verify that this document is valid.

Código para verificação: 74DF-171E-AEC0-78B5



Hash do Documento

360E391B1F252F3B629A50A57269229181F55C2E0A97394702041718422EDE8A

O(s) nome(s) indicado(s) para assinatura, bem como seu(s) status em 23/09/2021 é(são) :

- Paulo Sérgio de Paula Vargas (Reitor UFES) - 526.372.397-00 em 23/09/2021 17:40 UTC-03:00

Tipo: Assinatura Eletrônica

Identificação: Por email: reitor@ufes.br

Evidências

Client Timestamp Thu Sep 23 2021 17:40:49 GMT-0300 (Horário Padrão de Brasília)

Geolocation Location not shared by user.

IP 177.40.219.222

Hash Evidências:

25BD60492ED6357C707083AD929EE7B4E9110248E364D4F5E37AB5A264758D98

- Jorge Leonid Aching Samatelo (Testemunha) - 059.930.517-79 em 15/06/2021 15:34 UTC-03:00

Tipo: Assinatura Eletrônica

Identificação: Por email: jorge.samatelo@ufes.br

Evidências

Client Timestamp Tue Jun 15 2021 15:34:52 GMT-0300 (Horário Padrão de Brasília)

Geolocation Latitude: -20.2794336 Longitude: -40.2986108 Accuracy: 20

IP 177.157.219.41

Hash Evidências:

20B9F5249001A7D032C009EB145F3CE6AD8D2E2653056ABA57A5E02735B3EDF1

- ☑ André Resende Soares (Signatário) - 043.677.957-99 em 07/06/2021 17:17 UTC-03:00

Tipo: Assinatura Eletrônica

Identificação: Por email: Andre.Soares@vale.com

Evidências

Client Timestamp Mon Jun 07 2021 17:17:48 GMT-0300 (Horário Padrão de Brasília)

Geolocation Latitude: -20.222265 Longitude: -40.3186 Accuracy: 500

IP 189.14.205.251

Hash Evidências:

EA37AE73B9F1C1384C99543AECFE6367F0F1DFEE007815413AA190C19C04A9B7

- ☑ Rafael Gaier Gomes (Signatário) - 054.073.197-80 em 07/06/2021 17:01 UTC-03:00

Tipo: Assinatura Eletrônica

Identificação: Por email: Rafael.Gaier@vale.com

Evidências

Client Timestamp Mon Jun 07 2021 17:01:55 GMT-0300 (Horário Padrão de Brasília)

Geolocation Latitude: -20.4753988 Longitude: -45.1183088 Accuracy: 130

IP 152.238.78.5

Hash Evidências:

B672CA06F85EF70312668AA8EB6AB50B0F7F036C6600BC473B875C3C91328227

- ☑ Henrique Rieveres Borges de Andrade (Testemunha) - 884.547.897-15 em 07/06/2021 16:33 UTC-03:00

Tipo: Assinatura Eletrônica

Identificação: Por email: Henrique.Andrade@vale.com

Evidências

Client Timestamp Mon Jun 07 2021 16:33:15 GMT-0300 (Horário Padrão de Brasília)

Geolocation Latitude: -20.294857963271497 Longitude: -40.29991415751547 Accuracy: 168

IP 187.36.165.99

Hash Evidências:

D7A360C8EEFC8E456765D397D0B6A3C924270A327C3F154A242AED78938CA475

- ☑ Armando Biondo Filho (Superintendente FEST) - 376.717.407-30 em 07/06/2021 15:55 UTC-03:00

Tipo: Assinatura Eletrônica

Identificação: Por email: armando.biondo@fest.org.br

Evidências

Client Timestamp Mon Jun 07 2021 15:55:34 GMT-0300 (Horário Padrão de Brasília)

Geolocation Latitude: -20.2731 Longitude: -40.3044 Accuracy: 1090

IP 200.137.67.50

Hash Evidências:

1600C8B6722732B4AB905C803994049B44FC7A7956A7802976F4B312CB1A4264

