

CURSO DE SISTEMAS DE INFORMAÇÃO – TCC ACADÊMICO	
( X ) PRÉ-PROJETO    (   ) PROJETO	ANO/SEMESTRE: 2020/2

## YOUTUBORDERLESS: UNINDO LINGUAGENS E CULTURAS

Hugo Marcel Larsen

Prof.<sup>a</sup> Simone Erbs da Costa – Orientadora

### 1 INTRODUÇÃO

Em 11 de março de 2020, a Organização Mundial da Saúde (OMS) declarou que o novo Corona Vírus (nCoV-2019 ou Sars-cov-2), causador da doença Covid-19, se tornou uma pandemia (BARROZO *et al.* 2020). Para monitorar o surto atual, medidas robustas foram tomadas em todo o mundo para reduzir a transmissão do CORONA Vírus Disease 2019 (COVID-19), especialmente proibindo voos internacionais, ~~eriação-criando de~~ lockdowns em áreas vulneráveis, ~~e aplicando o~~ distanciamento social ~~etc.~~ (HAMID *et al.*, 2020).

Segundo Deslandes e Coutinho (2020), durante o isolamento social, o acesso à internet foi crescendo exponencialmente, sendo o principal meio de interatividade entre as pessoas. Com base na diminuição social entre as pessoas e suas drásticas mudanças de cotidiano, espera-se que o uso da internet sirva de normalizador de situações que antes não eram possíveis off-line e não agravem os efeitos colaterais que o isolamento social pode causar, como: depressão, ansiedade, solidão, violências familiares e suicídios (DESLANDES; COUTINHO, 2020).

Nesse sentido, o Youtube é amplamente utilizado como suporte pedagógico por uma quantidade significativa de internautas brasileiros conforme pesquisas, sendo a plataforma mais acessada para fins de conhecimento (18%), seguida da TV Globo (17%) e do Facebook (8%) (NAGUMO; TELES; SILVA, 2020). Xavier *et al.* (2019) também colocam o quanto o aparecimento de mídias e tecnologias, como o Youtube, ajudam tanto pessoas ouvintes como surdas, superando assim limites impostos no processo tradicional de ensino. Dessa forma, Xavier *et al.* (2019) ainda destacam a importância do uso do Youtube neste momento, devido a possibilidade de professores e educadores o usarem visando uma maior inclusão social quando comparadas as ferramentas tradicionais de educação.

Diante deste cenário, este trabalho propõe o desenvolvimento de um protótipo que auxiliará pessoas a entenderem transmissões ao vivo no Youtube, com auxílio de legendas. Conjectura-se que a construção deste protótipo quebre a barreira linguística entre conteúdos de interesse pessoal e auxilie em engajamentos interculturais e estudantis. Assim, objetivando também amenizar os efeitos colaterais citados causados durante o período da pandemia do COVID-19.

## 1.1 OBJETIVOS

O objetivo geral do trabalho é desenvolver um protótipo para auxiliar pessoas a entenderem transmissões ao vivo pelo Youtube com a ajuda de legendas em seu idioma. Sendo os objetivos específicos:

- a) disponibilizar uma aplicação em ambiente web e móvel;
- b) promover a aprendizagem por meio de transcrições de transmissão ao vivo pelo Youtube;
- c) promover a interação entre as diferentes culturas, amenizando os efeitos colaterais causados durante o período da pandemia do COVID-19;
- d) modelar a relação entre os requisitos da aplicação e práticas consolidadas no design de interface como as heurísticas de Nielsen, pelo método Relationship of M3C with User Requirements and Usability and Communicability Assessment in groupware (RURUCAg), possibilitando avaliar de maneira simples a usabilidade das aplicações disponibilizadas.

**Comentado [LPdAK1]:** Como fará essa promoção? EM que sentido?!

## 2 TRABALHOS CORRELATOS

Nesta seção serão descritos três projetos correlatos que apresentam características diretamente relacionadas ao trabalho proposto. A subseção 2.1 detalha o uso de um microcontrolador para transcrever a voz humana por meio de um dispositivo Android (UKO; DUBUKUMAH; AYOSUBOMI, 2019). A subseção 2.2 apresenta o desenvolvimento de uma aplicação de computador para traduzir e transcrever o idioma filipino para o idioma chinês (BAILON *et al.*, 2019). Por fim, a subseção 2.3 descreve a aplicação de computador KT-Speech-Crawler, na qual obtém dados de vídeos do Youtube para treinamento de redes neurais (LAKOMKIN *et al.*, 2018).

### 2.1 MICROCONTROLLER BASED SPEECH TO TEXT TRANSLATION SYSTEM

Uko, Dubukumah e Ayosubomi (2019) propuseram uma solução para resolver os problemas de uma comunicação efetiva em tempo real entre pessoas. No desenvolvimento do projeto foi utilizado o microcontrolador ATMEGA328p como recurso principal e “cérebro” do projeto; uma aplicação Android para captar a voz por um dispositivo móvel; um visor Liquid Crystal Display (LCD) para exibir o texto traduzido como a saída final para o usuário; um módulo Bluetooth para transmitir os dados da voz da aplicação Android; e outros componentes como provedores de energia (UKO; DUBUKUMAH; AYOSUBOMI, 2019).

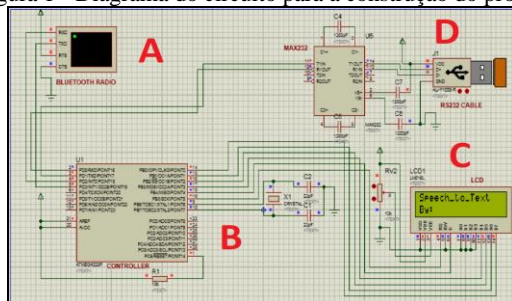
Segundo Uko, Dubukumah e Ayosubomi (2019), o trabalho teve como objetivo transcrever a fala humana em texto com um projeto de baixo custo. O intuito foi apresentar a metodologia, limitações, recomendações e aplicações para melhorar a comunicação humana com a integração de dispositivos eletrônicos, no qual não foi bem explorada devido complexidades como a captação de variações de sons para reconhecimento (UKO; DUBUKUMAH; AYOSUBOMI, 2019). Cabe destacar, algumas das características de Uko, Dubukumah e Ayosubomi (2019), como: reconhecimento de voz, transcrição de voz e disponibilização tanto na plataforma Android como por meio de componente.

Comentado [LPdAK2]: O que é componente?

O processo da transcrição de voz se inicia com a entrada da fala humana por meio do microfone no dispositivo Android. Duas aplicações são necessárias no dispositivo: o Adaptive Multi-Rate (AMR), um esquema de compressão de áudio usado para a codificação da voz; e o Google Voice Search utilizado como suporte para obter o vocabulário da linguagem a ser transcrita. Assim, os caracteres da linguagem reconhecida pelas aplicações são transferidos por comunicação serial por meio do módulo Bluetooth para o microcontrolador. Os dados dos caracteres precisam por fim serem convertidos de analógicos para digitais para serem exibidos nos visores (UKO; DUBUKUMAH; AYOSUBOMI, 2019).

O diagrama do circuito do projeto (Figura 1) demonstra como o aparelho desenvolvido tem seus módulos organizados. Na letra A pode ser visto o módulo de rádio Bluetooth que recebe as informações transferidas do dispositivo Android. A partir daí, o microcontrolador ATMEGA328p (letra B), “cérebro” do projeto, processa todas as funções programada no ambiente Arduino e compartilha suas informações com os outros módulos. Após cálculos e funções executadas, as letras C e D exibem o texto transcrito ao usuário, ou seja, a informação final. A letra C destaca o visor LCD incorporado no próprio dispositivo e a letra D a utilização do cabo RS 232 para exibição em visores maiores por exemplo (UKO; DUBUKUMAH; AYOSUBOMI, 2019).

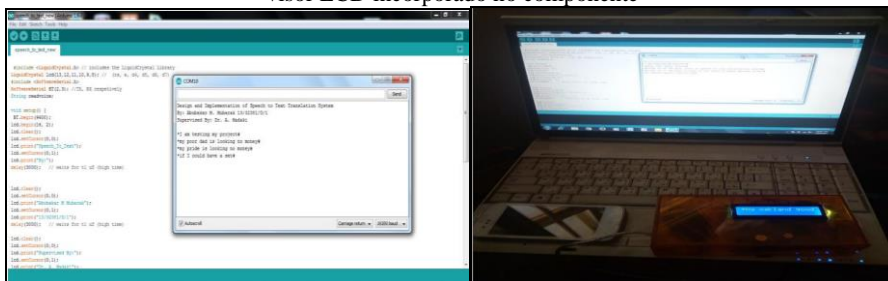
Figura 1 - Diagrama do circuito para a construção do projeto



Fonte: adaptada de Uko, Dubukumah e Ayosubomi (2019).

Uko, Dubukumah e Ayosubomi (2019) observam as limitações quanto ao uso do dispositivo desenvolvido, como: alcance da área de Bluetooth, tanto do dispositivo Android quanto do módulo Bluetooth; ruídos e barulhos que podem atrapalhar a captação genuína da voz; falta de mais telas para armazenamento do texto que já foi exibido transcrito; somente falas com um bom sotaque inglês eram traduzidas para texto; e o projeto foi limitado apenas com a língua inglesa. Ainda é possível observar a transcrição da voz humana sendo exibida de duas formas: no ambiente Arduino (Figura 2 (a)); e no próprio visor LCD incorporado no componente (Figura 2 (b)) (UKO; DUBUKUMAH; AYOSUBOMI, 2019).

Figura 2 – (a) Voz humana transcrita no ambiente Arduino e (b) Voz humana transcrita no visor LCD incorporado no componente



Fonte: adaptada de Uko, Dubukumah e Ayosubomi (2019).

Uko, Dubukumah e Ayosubomi (2019) colocam que por meio da realização desse projeto novas possibilidades e avanços na pesquisa de transcrição de texto são abordadas, como: fornecimento de um método renovado para ser aplicado na comunicação de dispositivos eletrônicos, como caixas eletrônicos, dispositivos multimídia ou robôs; aplicação em automação residencial e aplicações de segurança.

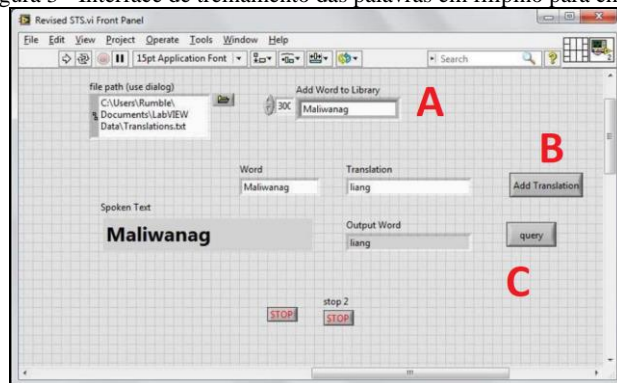
## 2.2 FILIPINO TO CHINESE SPEECH-TO-SPEECH TRANSLATOR USING NEURAL NETWORK WITH DATABASE SYSTEM

Bailon *et al.* (2019) tiveram como objetivo desenvolver uma aplicação que usa um algoritmo para traduzir a fala do idioma filipino para a fala no idioma chinês. O controlador principal da aplicação é uma Inteligência Artificial (IA) baseada em uma Rede Neural Artificial (RNA), auxiliada pela LabVIEW. As três etapas principais da aplicação são: reconhecimento da fala filipina para transcrição; tradução de texto filipino para chinês; reprodução falada do texto chinês (BAILON *et al.*, 2019). Cabe destacar, algumas das características de Bailon *et al.* (2019) como: reconhecimento de voz, transcrição de voz, tradução de texto e disponibilização para a plataforma desktop.

O fluxo do processo de estudo é separado em três estágios, sendo eles: de gravação; de recolhimento; e de testagem. No estágio de gravação são armazenadas 25 variações de fala de uma palavra. Isso é feito para 500 palavras frequentemente usadas no idioma filipino. Para o estágio de coleta, 500 palavras em filipino terão de corresponder a 500 palavras em chinês. Esse estágio também pode ser chamado de treinamento porque as variações de fala serão armazenadas na RNA e ela vai ser treinada para que reconheça as palavras específicas, independente do padrão de fala e ruídos externos (BAILON *et al.*, 2019).

Para ajudar no treinamento da RNA, a aplicação e a interface foram desenvolvidas para terem as seguintes funcionalidades (Figura 3): adicionar a palavra ao dicionário (letra A); inserir a tradução da palavra em filipino para chinês (letra B); utilizar um microfone para gravar a fala do usuário para a palavra correspondente (letra C). Nessa etapa, o papel do Windows Speech Recognition foi fundamental para deixar o reconhecimento da voz mais preciso, assim aumentando a acurácia em relação ao treinamento das palavras armazenadas na RNA (BAILON *et al.*, 2019).

Figura 3 - Interface de treinamento das palavras em filipino para chinês



Fonte: adaptada de Bailon et al. (2019).

Os resultados apresentados por Bailon *et al.* (2019) mostram lacunas em algumas palavras que não obtiveram uma taxa de sucesso alta na tradução porque possuem pronúncias parecidas. Entretanto, Bailon *et al.* (2019) observaram que a maioria das palavras teve baixo índice de rejeição, algumas possuindo a pontuação perfeita. Bailon *et al.* (2019) ainda colocam que para melhorar a taxa de sucesso das traduções, principalmente nas palavras com pontuação baixa, treinamentos com inúmeras iterações foram realizados. Com isso a taxa de sucesso, durante a etapa de treinamento, foi aumentada significativamente conforme os dados da Tabela 1. Entretanto, durante a fase de teste, realizada com 25 pessoas, a aplicação teve uma taxa de sucesso de 91,59% com uma taxa de rejeição de 8,41% (BAILON *et al.*, 2019).

Tabela 1 - Taxa de sucesso das palavras (Número de iterações/Percentual de sucesso)

# of Iterations	Success Rate %
25	86.8
37	88.2
49	89.6
61	91.8
73	92.0
85	93.4

Fonte: adaptada de Bailon et al. (2019).

Bailon *et al.* (2019) trazem recomendações propostas pelos pesquisadores da pesquisa, ~~são sendo~~ elas: mais variações de voz por palavra, que conforme comprovado, aumentam a eficácia de reconhecimento de palavras, juntamente com o número de iterações de treinamento; sugestão de aumentar a eficiência do reconhecimento da palavra de entrada para torná-la em torno de 97 a 99 por cento. Um por cento destinado a variáveis externas incontroláveis; aumento de duas a três mil palavras, englobando mais do vocabulário, gírias e expressões; e, ao longo prazo, uma aplicação expandida da RNA para reconhecimento de frases e sentenças, chegando assim a uma gramática correta e também mais próximo do objetivo de ter um tradutor totalmente controlado por computador (BAILON *et al.*, 2019).

### 2.3 KT-SPEECH-CRAWLER: AUTOMATIC DATASET CONSTRUCTION FOR SPEECH RECOGNITION FROM YOUTUBE VIDEOS

Para melhorar aplicações de reconhecimento de voz, Lakomkin *et al.* (2018) criaram o KT-Speech-Crawler, desenvolvido na linguagem Python, que tem o objetivo de recuperar informações de vídeos do Youtube. Essas informações são usadas para treinamento de aplicações neurais de reconhecimento de fala. A YouTube Search Application Programming Interface (API) foi a principal tecnologia utilizada para recuperar os vídeos do Youtube e suas legendas ocultas (LAKOMKIN *et al.*, 2018). Cabe destacar, algumas das características de Lakomkin *et al.* (2018), como: recuperar informações, reconhecimento de voz, transcrição de voz e disponibilização para a plataforma desktop.

O *crawler* pode obter cerca de 150 horas de fala transcrita em um dia, contendo uma taxa de erro de 3,5% no reconhecimento. As falas coletadas são gravadas em diversas condições, incluindo ruídos, músicas, má qualidade da voz por estar longe do microfone e diferentes sotaques e entonações. Foi observado uma redução de 40% na taxa de erros de palavras do conjunto de dados do Wall Street Journal apenas integrando 200 horas de amostras coletadas pelos vídeos no conjunto de treinamento (LAKOMKIN *et al.*, 2018).

As etapas principais executadas pelo KT-Speech-Crawler estão demonstradas na Figura 4 (a): processo de pesquisa dos vídeos pela API do Youtube, com palavras chaves da língua inglesa; ~~na letra B os filtros para aplicação tanto no áudio quanto nas legendas ocultas. São~~

são removidos por exemplo músicas de fundo e textos indesejados; na etapa de pós processamento (letra C) o programa tenta corrigir desalinhamentos imprecisos entre o áudio e as legendas do vídeo; a penúltima etapa (letra D) traz a transcrição do vídeo; e por fim, na letra E, a transcrição é enviada para a RNA ser treinada (LAKOMKIN *et al.*, 2018). Já a Figura 4 (b) traz a interface web desenvolvida por Lakomkin *et al.* (2018) para analisar as amostras retiradas e suas transcrições. O *web-service* desenvolvido exibe oito exemplos aleatórios e sua correspondente transcrição, permitindo carregar mais se necessário. Também foi integrada uma funcionalidade para que o usuário confirme se a transcrição está correta ou se não corresponde ao que foi falado no áudio. Lakomkin *et al.* (2018) estimam uma taxa de erro de 3,5% para um total de 600 amostras aleatórias.

Figura 4 – (a) Etapas do processo executado pelo KT-Speech-Crawler e (b) Interface web para análise das amostras e suas transcrições



Fonte: adaptada de Lakomkin et al. (2018).

Na demonstração de resultados, Lakomkin *et al.* (2018) provaram que adicionando amostras capturadas do Youtube aumentaram positivamente a performance do reconhecimento de voz dos dois conjuntos de dados, Wall Street Journal e TED-LIUM v2. Foram adicionadas 200 horas de amostras do Youtube (108.617 dicções) no conjunto de treinamento do Wall Street Journal e a taxa de erro de palavras foi melhorada de 34,2% para 15,8%. Um resultado similar foi visto no conjunto de dados TED-LIUM v2 ao adicionar 300 horas de amostras do Youtube, melhorando a taxa de erro de caracteres de 10,4% para 8,2% (LAKOMKIN *et al.*, 2018).

### 3 PROPOSTA DO PROTÓTIPO

Nesta seção serão apresentadas as justificativas para a realização do trabalho proposto (subseção 3.1), bem como os requisitos principais (subseção 3.2) e finalizando com a metodologia e o cronograma planejado para o desenvolvimento do trabalho (subseção 3.3).

**Comentado [LPdAK3]:** É isso mesmo? Tinha um .São O “.” estava errado?  
Estava confuso antes, mas não sei se assim está ok

**Comentado [LPdAK4]:** Sugiro dividir em frases;

### 3.1 JUSTIFICATIVA

Nas seções 1 e 2 foram evidenciadas a relevância do tema proposto. No Quadro 1 é apresentado um comparativo entre os trabalhos correlatos, de modo que as linhas representam as características e as colunas os trabalhos relacionados.

Quadro 1 - Comparativo entre os correlatos

Características	Correlatos	Uko, Dubukumah e Ayosubomi (2019)	Bailon <i>et al.</i> (2019)	Lakomkin <i>et al.</i> (2018)
Recuperar informações de vídeos do Youtube		X	X	✓
Reconhecimento de voz		✓	✓	✓
Transcrição de voz		✓	✓	✓
Tradução de texto		X	✓	X
Plataforma		Componente	Desktop	Desktop

Fonte: elaborado pelo autor.

Comentado [LPdAK5]: O que é plataforma "componente"?

Conforme demonstrado no Quadro 1, percebe-se que Uko, Dubukumah e Ayosubomi (2019), Bailon *et al.* (2019) e Lakomkin *et al.* (2018) têm finalidades semelhantes no que se diz respeito a reconhecimento e transcrição de voz. Ambas as características foram identificadas nos três correlatos como essenciais para alcançar seus objetivos, uma vez que são importantes para a operacionalização das demais características. Lakomkin *et al.* (2018) se destacam por apresentarem a característica relacionada a recuperar informações de vídeos do Youtube, dizendo respeito a recuperar dados da voz e outras informações.

O reconhecimento de voz é uma característica que foi vista em todos os trabalhos correlatos. Em Uko, Dubukumah e Ayosubomi (2019) a voz era recuperada por um dispositivo Android e transmitida pelo módulo Bluetooth para o componente desenvolvido. Bailon *et al.* (2019), para treinarem a RNA, utilizavam-se do Windows Speech Recognition para reconhecer e melhorar o reconhecimento de voz por parte dos participantes. Já Lakomkin *et al.* (2018) fizeram uso das informações de vídeos do Youtube para obter a voz e treinar redes neurais.

Assim como a característica de reconhecimento de voz, a transcrição de voz também é fundamental nos três correlatos. Uko, Dubukumah e Ayosubomi (2019) precisam dessa característica para que o usuário saiba o que foi dito por outra pessoa. Bailon *et al.* (2019) necessitam transcrever a voz filipina para que seja possível traduzir o texto para chinês e manter a comunicação desses dois idiomas. Lakomkin *et al.* (2018) fizeram uso dessa característica para treinar redes neurais e conferir os resultados obtidos das amostras.

A característica de tradução de texto foi vista somente em Bailon *et al.* (2019). Ela se assemelha a este protótipo porque a pessoas que não falam o mesmo idioma possam entender o conteúdo da transmissão ao vivo. Bailon *et al.* (2019) têm seu principal objetivo voltado



para essa característica, na qual a comunicação entre filipino e chinês possa ser compreendida com resultados satisfatórios.

Com base nestas características e no Quadro 1 apresentado, nota-se que o trabalho proposto se torna relevante ao buscar ajudar pessoas a entenderem transmissões ao vivo que não são de seu idioma. Além disso, a proposta visa auxiliar em outros aspectos como a integração de novos conteúdos na vida das pessoas para que possam aprofundar estudos e até mesmo se divertir ao entenderem o que está acontecendo em transmissões ao vivo. Já como contribuição acadêmica será utilizado o Método RURUCaG, que poderá ser utilizado em trabalhos que queiram modelar a relação entre os requisitos da aplicação e práticas consolidadas no design de interface como as heurísticas de Nielsen, bem como avaliar a usabilidade e a experiência de usuários em aplicações na área da computação. Além disso, traz a contribuição acadêmica buscando encontrar interfaces amigáveis para serem aplicadas no protótipo, permitindo que os usuários finais se sintam mais confortáveis e foquem em seus objetivos. Como contribuição tecnológica pode-se destacar o desenvolvimento de uma aplicação web e aplicação móvel, com a tecnologia Progressive Web Application (PWA), além de fazer uso de um Sistema Gerenciador de Banco de Dados (SGBD) Not Only Structured Query Language (NoSQL).

**Comentado [LPdAK6]:** Rever com a sua orientadora, pois esse texto está idêntico na proposta de seus outros orientandos. No meu ponto de vista, cada um deve escrever com suas palavras a sua percepção, já que é a justificativa deste trabalho. Se quiserem escrever igual, colocar entre "" e a referência de onde tiraram isso.

### 3.2 REQUISITOS PRINCIPAIS DO PROBLEMA A SER TRABALHADO

Nesta subseção serão abordados os principais Requisitos Funcionais (RF), assim como os principais Requisitos Não Funcionais (RNF), conforme Quadro 2.

Quadro 2 – Principais Requisitos Funcionais e Não Funcionais

A aplicação deverá:	Tipo
permitir ao usuário acesso por meio de uma conta Google.	RF
permitir ao usuário consultar uma lista de vídeos ao vivo do Youtube conforme as inscrições do usuário autenticado com a conta Google.	RF
permitir ao usuário visualizar um vídeo ao vivo.	RF
permitir ao usuário configurar a transcrição para o vídeo ao vivo sendo visualizado.	RF
utilizar serviços Google Cloud para transcrever áudio e traduzir essa transcrição.	RNF
utilizar APIs para recuperar informações de vídeos ao vivo do Youtube.	RNF
implementar a aplicação web utilizando o <i>framework</i> Angular.	RNF
implementar a aplicação web seguindo as diretrizes do Material Design (MD).	RNF
consultar serviços na parte do servidor desenvolvido em NodeJS.	RNF
implementar uma aplicação móvel com a tecnologia PWA.	RNF
implementar uma aplicação móvel com a ajuda do ambiente de desenvolvimento Visual Studio Code.	RNF
utilizar a plataforma Firebase (Google) para usar serviços de hospedagem e validar acessos de contas do Google.	RNF
utilizar a plataforma Firebase como SGBD NoSQL para armazenar informações dos usuários.	RNF
utilizar o Método RURUCaG para modelar os requisitos da aplicação com as heurísticas de Nielsen.	RNF
utilizar o Método RURUCaG para avaliar a usabilidade e a experiência de uso da aplicação.	RNF

Fonte: elaborado pelo autor.

**Comentado [LPdAK7]:** Remover “.” das frases.

**Comentado [LPdAK8]:** Acho que essa primeira é um RNF

### 3.3 METODOLOGIA

A metodologia desta proposta está elaborada em seis etapas e composta pelos instrumentos metodológicos contidos no Quadro 3, juntamente com os períodos relacionados.

Quadro 3 – Cronograma com as etapas e metodologias

Etapas	Quinzenas		2021									
			fev.		mar.		abr.		maio		jun.	
	1	2	1	2	1	2	1	2	1	2	1	2
<b>Pesquisa na literatura:</b> realizar uma revisão mais aprofundada da literatura sobre os assuntos citados na revisão bibliográfica e trabalhos correlatos.												
<b>Refinamento dos requisitos:</b> reavaliar os requisitos funcionais e não funcionais já definidos e, se necessário.												
<b>Especificação e análise:</b> formalizar as funcionalidades da ferramenta por meio de casos de uso e diagramas de atividade da Unified Modeling Language (UML), utilizando a ferramenta Astah Community.												
<b>Implementação da aplicação web:</b> implementar uma aplicação web para os usuários finais utilizando como base a linguagem de programação JavaScript e Visual Studio Code como ambiente de desenvolvimento.												
<b>Implementação da aplicação móvel:</b> implementar uma aplicação móvel com a tecnologia PWA para exibir o conteúdo da aplicação web e Visual Studio Code como ambiente de desenvolvimento.												
<b>Verificação e validação:</b> validar a usabilidade da solução pelo Método RURUCAg.												

Fonte: elaborado pelo autor.

## 4 REVISÃO BIBLIOGRÁFICA

Conforme os dados da Pesquisa Nacional por Amostra de Domicílios Contínua (PNAD) realizada no ano de 2016, uma das atividades principais do brasileiro na internet é assistir vídeos on-line no celular (IBGE, 2018). Segundo Nagumo, Teles e Silva (2020), a pesquisa Video Viewers realizada com três mil pessoas pelo Instituto Provokers, em 2018, mostrou que nove em cada dez brasileiros usam o Youtube para aprendizado. Com base no crescimento de acessos ao Youtube pelo mundo foi criada a Youtube Data API para simplificar e unificar a forma de capturar e enviar informações relacionadas a vídeos e canais no Youtube (GOOGLE, 2020).

O Google Tradutor é possivelmente o serviço de tradução mais utilizado, e em 2016, mais de 500 milhões de pessoas estavam traduzindo 100 bilhões de palavras por dia em 103 idiomas, sendo o Brasil com maior número de usuários (BROOKS, 2016). Existem diferentes tecnologias que podem alcançar a implementação de fala e uma delas é a Sistemas de Rede Neural (SRN) (BAILON *et al.*, 2019). Uma Rede Neural Recorrente (RNR) é uma arquitetura de RNA que contém conexões cíclicas, que permitem aprender a dinâmica temporal de dados sequenciais (MURAD; PYUN, 2017).

Para se ter uma maior interação e feedback com o usuário, o uso de interfaces amigáveis é indispensável em uma aplicação desenvolvida, independente da tecnologia

utilizada (COSTA, 2018). Segundo Google (2019), o Material Design (MD) foi criado para que o usuário tenha menos esforço ao utilizar uma aplicação e **consiga focar no que ele precisa**. Torres *et al.* (2018) citam que o MD foi gerado para manter uma identidade visual em diferentes aplicações como em sites e no sistema operacional **Android**. De acordo com Costa (2018), o método RURUCAg foi criado para avaliar a interface de aplicações **com base nas** heurísticas de Nielsen e assim reconhecer incidentes com relação a usabilidade, visando assim melhorar a interface e funcionalidades da aplicação para o usuário final.

Por fim, cabe destacar que **com** o aumento de projetos para aplicações móveis também surgiu a dificuldade de reutilização de código para aplicações existentes. O PWA foi criado para, de forma híbrida, unificar e minimizar implementações entre aplicações web e móveis. Para Sharma *et al.* (2019), utilizar o PWA implica em um baixo tempo de desenvolvimento e de implantação da aplicação, além de permitir que a empresa não contrate funcionários especializados em diferentes plataformas. As funcionalidades do PWA incluem: funcionamento off-line; sincronização em segundo plano; e instalação da aplicação web em plataformas móveis. Além de não precisar de uma loja de aplicativos para que a aplicação seja distribuída (BIØRN-HANSEN *et al.*, 2017).

#### REFERÊNCIAS

BAILON, Mark R. M.; DE SILVA, Mark A. C.; LAPUZ, Reyann J. P.; TINIO, John L. A.; YU, Ted B. K.; GUSTILO, Reggie C. **Filipino to chinese speech-to-speech translator using neural network with database system**. International Journal of Emerging Trends in Engineering Research, [S. l.], v. 7, n. 9, p. 276–282, 2019.

BARROZO, Ligia V.; SERAFIM, Mirela B.; MORAES, Sara L. de; MANSUR, Giselle. **MONITORAMENTO ESPAÇO-TEMPORAL DAS ÁREAS DE ALTO RISCO DE COVID-19 NOS MUNICÍPIOS DO BRASIL**. Hygeia - Revista Brasileira de Geografia Médica e da Saúde, [S. l.], p. 417–425, 2020.

BIØRN-HANSEN, Andreas; MAJCHRZAK, Tim A.; GRØNLI, Tor M. **Progressive web apps: The possibleweb-native unifier for mobile development**. In: WEBIST 2017 - PROCEEDINGS OF THE 13TH INTERNATIONAL CONFERENCE ON WEB INFORMATION SYSTEMS AND TECHNOLOGIES 2017, Anais [...]. [s.l.: s.n.].

BROOKS, Richard. **11 Google Translate Facts You Should Know**. K International, 02 mai. 2016. Disponível em: <https://k-international.com/blog/google-translate-facts/>. Acesso em: 14 out. 2020.

COSTA, Simone Erbs da. **iLibras como facilitador na comunicação do surdo: Desenvolvimento de um recurso colaborativo de tecnologia assistiva**. 2018. 263 f. Dissertação (Mestrado) – Centro de Ciências Tecnológicas, Universidade do Estado de Santa Catarina, 2018.

**Comentado [LPdAK9]:** Parece muito coloquial.

**Comentado [LPdAK10]:** Acho que essas duas frases não se contextualizam. Dessa forma, falta um link entre.

**Comentado [LPdAK11]:** Referenciar

DESLANDES, Suely F.; COUTINHO, Tiago. **The intensive use of the internet by children and adolescents in the context of COVID-19 and the risks for self-inflicted violence.** *Ciência e Saúde Coletiva*, [S. l.], v. 25, p. 2479–2486, 2020.

GOOGLE. Material Design. **Introduction.** [S.l.], 2019. Disponível em: <https://material.io/design/introduction/>. Acesso em: 08 out. 2020.

GOOGLE. Youtube Data API. **Página inicial**, 2020. Disponível em: <https://developers.google.com/youtube/v3>. Acesso em: 08 out. 2020.

HAMID, Saima; MIR, Mohammad Y.; ROHELA, Gulab K. **Novel coronavirus disease (COVID-19): a pandemic (epidemiology, pathogenesis and potential therapeutics).** *New Microbes and New Infections*, 2020.

IBGE. **Acesso à internet e à televisão e posse de telefone móvel celular para uso pessoal: 2016.** Rio de Janeiro: IBGE, 2018.

LAKOMKIN, Egor; MAGG, Sven; WEBER, Cornelius; WERMTER, Stefan. **KT-Speech-Crawler: Automatic dataset construction for speech recognition from YouTube videos.** EMNLP 2018 - CONFERENCE ON EMPIRICAL METHODS IN NATURAL LANGUAGE PROCESSING: SYSTEM DEMONSTRATIONS, PROCEEDINGS 2018.

MURAD, Abdulmajid; PYUN, Jae Y. **Deep recurrent neural networks for human activity recognition.** *Sensors (Switzerland)*, [S. l.], v. 17, n. 11, 2017.

NAGUMO, Estevon; TELES, Lúcio F.; SILVA, Lucélia de A. **A utilização de vídeos do Youtube como suporte ao processo de aprendizagem (Using Youtube videos to support the learning process).** *Revista Eletrônica de Educação*, [S. l.], v. 14, p. 3757008, 2020.

SHARMA, Varsha; VERMA, Rajat; PATHAK, Vaishali; PALIWAL, Muskan; JAIN, Priya. **Progressive Web App (PWA) - One Stop Solution for All Application Development Across All Platforms.** *International Journal of Scientific Research in Computer Science, Engineering and Information Technology*, [S. l.], p. 1120–1122, 2019.

TORRES, Cecília V. et al. **DESENVOLVIMENTO MOBILE COM ENFOQUE ACESSÍVEL: O DESIGN NA MEDIAÇÃO DA INCLUSÃO.** *Human Factors in Design*, [S. l.], v. 7, n. 13, p. 85–101, 2018.

UKO, Victor S.; DUBUKUMAH, Gachada B.; AYOSUBOMI, Ibrahim K. **Microcontroller Based Speech to Text Translation System.** *European Journal of Engineering Research and Science*, [S. l.], v. 4, n. 12, p. 149–154, 2019.

XAVIER, Marcelle B.; CARVALHO, Francisco dos S.; CARVALHO, Mauro dos S.; MORAES, Juliana M. de. **Identidade Surda: Uma Análise dos Discursos de Professores Surdos e Ouvintes no Youtube / Deaf Identity: An Analysis of the Discourses of Deaf Teachers and Listeners on Youtube.** ID on line REVISTA DE PSICOLOGIA, [S. l.], v. 13, n. 45, 2019.

**ASSINATURAS**

(Atenção: todas as folhas devem estar rubricadas)

Assinatura do(a) Aluno(a): \_\_\_\_\_

Assinatura do(a) Orientador(a): \_\_\_\_\_

Assinatura do(a) Coorientador(a) (se houver): \_\_\_\_\_

Observações do orientador em relação a itens não atendidos do pré-projeto (se houver):

## FORMULÁRIO DE AVALIAÇÃO – PROFESSOR AVALIADOR

Acadêmico(a): \_\_\_\_\_

Avaliador(a): Luciana Pereira de Araújo Kohler

ASPECTOS AVALIADOS <sup>1</sup>		atende	atende parcialmente	não atende
ASPECTOS TÉCNICOS	1. INTRODUÇÃO O tema de pesquisa está devidamente contextualizado/delimitado?	X		
	O problema está claramente formulado?	X		
	1. OBJETIVOS O objetivo principal está claramente definido e é passível de ser alcançado?	X		
	Os objetivos específicos são coerentes com o objetivo principal?		X	
	2. TRABALHOS CORRELATOS São apresentados trabalhos correlatos, bem como descritas as principais funcionalidades e os pontos fortes e fracos?	X		
	3. JUSTIFICATIVA Foi apresentado e discutido um quadro relacionando os trabalhos correlatos e suas principais funcionalidades com a proposta apresentada?	X		
	São apresentados argumentos científicos, técnicos ou metodológicos que justificam a proposta?	X		
	São apresentadas as contribuições teóricas, práticas ou sociais que justificam a proposta?	X		
	4. REQUISITOS PRINCIPAIS DO PROBLEMA A SER TRABALHADO Os requisitos funcionais e não funcionais foram claramente descritos?	X		
	5. METODOLOGIA Foram relacionadas todas as etapas necessárias para o desenvolvimento do TCC?	X		
	Os métodos, recursos e o cronograma estão devidamente apresentados e são compatíveis com a metodologia proposta?	X		
	6. REVISÃO BIBLIOGRÁFICA (atenção para a diferença de conteúdo entre projeto e pré-projeto) Os assuntos apresentados são suficientes e têm relação com o tema do TCC?	X		
ASPECTOS METODOLÓGICOS	As referências contemplam adequadamente os assuntos abordados (são indicadas obras atualizadas e as mais importantes da área)?	X		
	7. LINGUAGEM USADA (redação) O texto completo é coerente e redigido corretamente em língua portuguesa, usando linguagem formal/científica?	X		
	A exposição do assunto é ordenada (as ideias estão bem encadeadas e a linguagem utilizada é clara)?	X		

### PARECER – PROFESSOR AVALIADOR: (PREENCHER APENAS NO PROJETO)

O projeto de TCC deverá ser revisado, isto é, necessita de complementação, se:

- qualquer um dos itens tiver resposta NÃO ATENDE;
- pelo menos 5 (cinco) tiverem resposta ATENDE PARCIALMENTE.

**PARECER:** ( ) APROVADO ( ) REPROVADO

Assinatura: \_\_\_\_\_ Data: \_\_\_\_\_

<sup>1</sup> Quando o avaliador marcar algum item como atende parcialmente ou não atende, deve obrigatoriamente indicar os motivos no texto, para que o aluno saiba o porquê da avaliação.

CURSO DE SISTEMAS DE INFORMAÇÃO – TCC ACADÊMICO	
( X ) PRÉ-PROJETO    (   ) PROJETO	ANO/SEMESTRE: 2020/2

## YOUTUBORDERLESS: UNINDO LINGUAGENS E CULTURAS

Hugo Marcel Larsen

Prof.<sup>a</sup> Simone Erbs da Costa – Orientadora

### 1 INTRODUÇÃO

Em 11 de março de 2020, a Organização Mundial da Saúde (OMS) declarou que o novo Corona Vírus (nCoV-2019 ou Sars-cov-2), causador da doença Covid-19, se tornou uma pandemia (BARROZO *et al.* 2020). Para monitorar o surto atual, medidas robustas foram tomadas em todo o mundo para reduzir a transmissão do Corona Vírus Disease 2019 (COVID-19), especialmente proibindo voos internacionais, criação de lockdowns em áreas vulneráveis, distanciamento social etc. (HAMID *et al.*, 2020).

Segundo Deslandes e Coutinho (2020), durante o isolamento social, o acesso à internet foi crescendo exponencialmente, sendo o principal meio de interatividade entre as pessoas. Com base na diminuição social entre as pessoas e suas drásticas mudanças de cotidiano, espera-se que o uso da internet sirva de normalizador de situações que antes não eram possíveis off-line e não agravem os efeitos colaterais que o isolamento social pode causar, como: depressão, ansiedade, solidão, violências familiares e suicídios (DESLANDES; COUTINHO, 2020).

Nesse sentido, o Youtube é amplamente utilizado como suporte pedagógico por uma quantidade significativa de internautas brasileiros conforme pesquisas, sendo a plataforma mais acessada para fins de conhecimento (18%), seguida da TV Globo (17%) e do Facebook (8%) (NAGUMO; TELES; SILVA, 2020). Xavier *et al.* (2019) também colocam o quanto o aparecimento de mídias e tecnologias, como o Youtube, ajudam tanto pessoas ouvintes como surdas, superando assim limites impostos no processo tradicional de ensino. Dessa forma, Xavier *et al.* (2019) ainda destacam a importância do uso do Youtube neste momento, devido a possibilidade de professores e educadores o usarem visando uma maior inclusão social quando comparadas as ferramentas tradicionais de educação.

Diante deste cenário, este trabalho propõe o desenvolvimento de um protótipo que auxiliará pessoas a entenderem transmissões ao vivo no Youtube, com auxílio de legendas. Conjectura-se que a construção deste protótipo quebre a barreira linguística entre conteúdos de interesse pessoal e auxilie em engajamentos interculturais e estudantis. Assim, objetivando também amenizar os efeitos colaterais citados causados durante o período da pandemia do COVID-19.

Comentado [AS1]: Não foram criados.

## 1.1 OBJETIVOS

O objetivo geral do trabalho é desenvolver um protótipo para auxiliar pessoas a entenderem transmissões ao vivo pelo Youtube com a ajuda de legendas em seu idioma. Sendo os objetivos específicos:

- a) disponibilizar uma aplicação em ambiente web e móvel;
- b) promover a aprendizagem por meio de transcrições de transmissão ao vivo pelo Youtube;
- c) promover a interação entre as diferentes culturas, amenizando os efeitos colaterais causados durante o período da pandemia do COVID-19;
- d) modelar a relação entre os requisitos da aplicação e práticas consolidadas no design de interface como as heurísticas de Nielsen, pelo método Relationship of M3C with User Requirements and Usability and Communicability Assessment in groupware (RURUCAG), possibilitando avaliar de maneira simples a usabilidade das aplicações disponibilizadas.

**Comentado [AS2]:** Desenvolver é metodologia e não objetivo.

**Comentado [AS3]:** De que forma será esta interação?

## 2 TRABALHOS CORRELATOS

Nesta seção serão descritos três projetos correlatos que apresentam características diretamente relacionadas ao trabalho proposto. A subseção 2.1 detalha o uso de um microcontrolador para transcrever a voz humana por meio de um dispositivo Android (UKO; DUBUKUMAH; AYOSUBOMI, 2019). A subseção 2.2 apresenta o desenvolvimento de uma aplicação de computador para traduzir transcrever o idioma filipino para o idioma chinês (BAILON *et al.*, 2019). Por fim, a subseção 2.3 descreve a aplicação de computador KT-Speech-Crawler, na qual obtém dados de vídeos do Youtube para treinamento de redes neurais (LAKOMKIN *et al.*, 2018).

### 2.1 MICROCONTROLLER BASED SPEECH TO TEXT TRANSLATION SYSTEM

Uko, Dubukumah e Ayosubomi (2019) propuseram uma solução para resolver os problemas de uma comunicação efetiva em tempo real entre pessoas. No desenvolvimento do projeto foi utilizado o microcontrolador ATMEGA328p como recurso principal e “cérebro” do projeto; uma aplicação Android para captar a voz por um dispositivo móvel; um visor Liquid Crystal Display (LCD) para exibir o texto traduzido como a saída final para o usuário; um módulo Bluetooth para transmitir os dados da voz da aplicação Android; e outros componentes como provedores de energia (UKO; DUBUKUMAH; AYOSUBOMI, 2019).



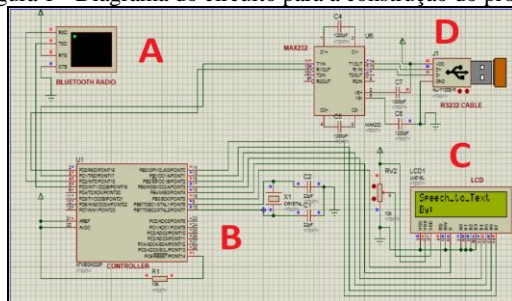
Segundo Uko, Dubukumah e Ayosubomi (2019), o trabalho teve como objetivo transcrever a fala humana em texto com um projeto de baixo custo. O intuito foi apresentar a metodologia, limitações, recomendações e aplicações para melhorar a comunicação humana com a integração de dispositivos eletrônicos, no qual não foi bem explorada devido complexidades como a captação de variações de sons para reconhecimento (UKO; DUBUKUMAH; AYOSUBOMI, 2019). Cabe destacar, algumas das características de Uko, Dubukumah e Ayosubomi (2019), como: reconhecimento de voz, transcrição de voz e disponibilização tanto na plataforma Android como por meio de componente.

Comentado [AS4]: O que é?

O processo da transcrição de voz se inicia com a entrada da fala humana por meio do microfone no dispositivo Android. Duas aplicações são necessárias no dispositivo: o Adaptive Multi-Rate (AMR), um esquema de compressão de áudio usado para a codificação da voz; e o Google Voice Search utilizado como suporte para obter o vocabulário da linguagem a ser transcrita. Assim, os caracteres da linguagem reconhecida pelas aplicações são transferidos por comunicação serial por meio do módulo Bluetooth para o microcontrolador. Os dados dos caracteres precisam por fim serem convertidos de analógicos para digitais para serem exibidos nos visores (UKO; DUBUKUMAH; AYOSUBOMI, 2019).

O diagrama do circuito do projeto (Figura 1) demonstra como o aparelho desenvolvido tem seus módulos organizados. Na letra A pode ser visto o módulo de rádio Bluetooth que recebe as informações transferidas do dispositivo Android. A partir daí, o microcontrolador ATMEGA328p (letra B), “cérebro” do projeto, processa todas as funções programadas no ambiente Arduino e compartilha suas informações com os outros módulos. Após cálculos e funções executadas, as letras C e D exibem o texto transcrito ao usuário, ou seja, a informação final. A letra C destaca o visor LCD incorporado no próprio dispositivo e a letra D a utilização do cabo RS 232 para exibição em visores maiores por exemplo (UKO; DUBUKUMAH; AYOSUBOMI, 2019).

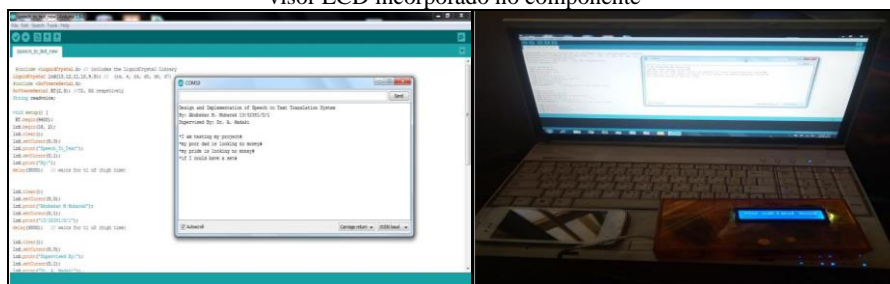
Figura 1 - Diagrama do circuito para a construção do projeto



Fonte: adaptada de Uko, Dubukumah e Ayosubomi (2019).

Uko, Dubukumah e Ayosubomi (2019) observam as limitações quanto ao uso do dispositivo desenvolvido, como: alcance da área de Bluetooth, tanto do dispositivo Android quanto do módulo Bluetooth; ruídos e barulhos que podem atrapalhar a captação genuína da voz; falta de mais telas para armazenamento do texto que já foi exibido transcrito; somente falas com um bom sotaque inglês eram traduzidas para texto; e o projeto foi limitado apenas com a língua inglesa. Ainda é possível observar a transcrição da voz humana sendo exibida de duas formas: no ambiente Arduino (Figura 2 (a)); e no próprio visor LCD incorporado no componente (Figura 2 (b)) (UKO; DUBUKUMAH; AYOSUBOMI, 2019).

Figura 2 – (a) Voz humana transcrita no ambiente Arduino e (b) Voz humana transcrita no visor LCD incorporado no componente



Fonte: adaptada de Uko, Dubukumah e Ayosubomi (2019).

Uko, Dubukumah e Ayosubomi (2019) colocam que por meio da realização desse projeto novas possibilidades e avanços na pesquisa de transcrição de texto são abordadas, como: fornecimento de um método renovado para ser aplicado na comunicação de dispositivos eletrônicos, como caixas eletrônicas, dispositivos multimídia ou robôs; aplicação em automação residencial e aplicações de segurança.

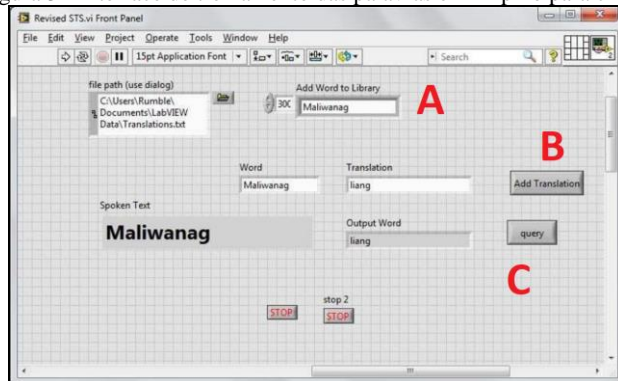
## 2.2 FILIPINO TO CHINESE SPEECH-TO-SPEECH TRANSLATOR USING NEURAL NETWORK WITH DATABASE SYSTEM

Bailon *et al.* (2019) tiveram como objetivo desenvolver uma aplicação que usa um algoritmo para traduzir a fala do idioma filipino para a fala no idioma chinês. O controlador principal da aplicação é uma Inteligência Artificial (IA) baseada em uma Rede Neural Artificial (RNA), auxiliada pela LabVIEW. As três etapas principais da aplicação são: reconhecimento da fala filipina para transcrição; tradução de texto filipino para chinês; reprodução falada do texto chinês (BAILON *et al.*, 2019). Cabe destacar, algumas das características de Bailon *et al.* (2019) como: reconhecimento de voz, transcrição de voz, tradução de texto e disponibilização para a plataforma desktop.

O fluxo do processo de estudo é separado em três estágios, sendo eles: de gravação; de recolhimento; e de testagem. No estágio de gravação são armazenadas 25 variações de fala de uma palavra. Isso é feito para 500 palavras frequentemente usadas no idioma filipino. Para o estágio de coleta, 500 palavras em filipino terão de corresponder a 500 palavras em chinês. Esse estágio também pode ser chamado de treinamento porque as variações de fala serão armazenadas na RNA e ela vai ser treinada para que reconheça as palavras específicas, independente do padrão de fala e ruídos externos (BAILON *et al.*, 2019).

Para ajudar no treinamento da RNA, a aplicação e a interface foram desenvolvidas para terem as seguintes funcionalidades (Figura 3): adicionar a palavra ao dicionário (letra A); inserir a tradução da palavra em filipino para chinês (letra B); utilizar um microfone para gravar a fala do usuário para a palavra correspondente (letra C). Nessa etapa o papel do Windows Speech Recognition foi fundamental para deixar o reconhecimento da voz mais preciso, assim aumentando a acurácia em relação ao treinamento das palavras armazenadas na RNA (BAILON *et al.*, 2019).

Figura 3 - Interface de treinamento das palavras em filipino para chinês



Fonte: adaptada de Bailon et al. (2019).

Os resultados apresentados por Bailon *et al.* (2019) mostram lacunas em algumas palavras que não obtiveram uma taxa de sucesso alta na tradução porque possuem pronúncias parecidas. Entretanto, Bailon *et al.* (2019) observaram que a maioria das palavras teve baixo índice de rejeição, algumas possuindo a pontuação perfeita. Bailon *et al.* (2019) ainda colocam que para melhorar a taxa de sucesso das traduções, principalmente nas palavras com pontuação baixa, treinamentos com inúmeras iterações foram realizados. Com isso, a taxa de sucesso, durante a etapa de treinamento, foi aumentada significativamente conforme os dados da Tabela 1. Entretanto, durante a fase de teste, realizada com 25 pessoas, a aplicação teve uma taxa de sucesso de 91,59% com uma taxa de rejeição de 8,41% (BAILON *et al.*, 2019).

Tabela 1 - Taxa de sucesso das palavras (Número de iterações/Percentual de sucesso)

# of Iterations	Success Rate %
25	86.8
37	88.2
49	89.6
61	91.8
73	92.0
85	93.4

Fonte: adaptada de Bailon et al. (2019).

Comentado [AS5]: Você pode colocar isso diretamente na tabela

Bailon *et al.* (2019) trazem recomendações propostas pelos pesquisadores da pesquisa, são elas: mais variações de voz por palavra, que conforme comprovado, aumentam a eficácia de reconhecimento de palavras, juntamente com o número de iterações de treinamento; sugestão de aumentar a eficiência do reconhecimento da palavra de entrada para torná-la em torno de 97 a 99 por cento. Um por cento destinado a variáveis externas incontroláveis; aumento de duas a três mil palavras, englobando mais do vocabulário, gírias e expressões; e, ao longo prazo, uma aplicação expandida da RNA para reconhecimento de frases e sentenças, chegando assim a uma gramática correta e também mais próximo do objetivo de ter um tradutor totalmente controlado por computador (BAILON *et al.*, 2019).

Comentado [AS6]: Esta frase é continuação da frase anterior?

### 2.3 KT-SPEECH-CRAWLER: AUTOMATIC DATASET CONSTRUCTION FOR SPEECH RECOGNITION FROM YOUTUBE VIDEOS

Para melhorar aplicações de reconhecimento de voz, Lakomkin *et al.* (2018) criaram o KT-Speech-Crawler, desenvolvido na linguagem Python, que tem o objetivo de recuperar informações de vídeos do Youtube. Essas informações são usadas para treinamento de aplicações neurais de reconhecimento de fala. A YouTube Search Application Programming Interface (API) foi a principal tecnologia utilizada para recuperar os vídeos do Youtube e suas legendas ocultas (LAKOMKIN *et al.*, 2018). Cabe destacar, algumas das características de Lakomkin *et al.* (2018), como: recuperar informações, reconhecimento de voz, transcrição de voz e disponibilização para a plataforma desktop.

Comentado [AS7]: O que é isso? São RNAs?

O *crawler* pode obter cerca de 150 horas de fala transcrita em um dia, contendo uma taxa de erro de 3,5% no reconhecimento. As falas coletadas são gravadas em diversas condições, incluindo ruídos, músicas, má qualidade da voz por estar longe do microfone e diferentes sotaques e entonações. Foi observado uma redução de 40% na taxa de erros de palavras do conjunto de dados do Wall Street Journal apenas integrando 200 horas de amostras coletadas pelos vídeos no conjunto de treinamento (LAKOMKIN *et al.*, 2018).

As etapas principais executadas pelo KT-Speech-Crawler estão demonstradas na Figura 4 (a): processo de pesquisa dos vídeos pela API do Youtube, com palavras chaves da língua inglesa; na letra B os filtros para aplicação tanto no áudio quanto nas legendas ocultas. São

removidos por exemplo músicas de fundo e textos indesejados; na etapa de pós processamento (letra C) o programa tenta corrigir desalinhamentos imprecisos entre o áudio e as legendas do vídeo; a penúltima etapa (letra D) traz a transcrição do vídeo; e por fim, na letra E, a transcrição é enviada para a RNA ser treinada (LAKOMKIN *et al.*, 2018). Já a Figura 4 (b) traz a interface web desenvolvida por Lakomkin *et al.* (2018) para analisar as amostras retiradas e suas transcrições. O *web-service* desenvolvido exibe oito exemplos aleatórios e sua correspondente transcrição, permitindo carregar mais se necessário. Também foi integrada uma funcionalidade para que o usuário confirme se a transcrição está correta ou se não corresponde ao que foi falado no áudio. Lakomkin *et al.* (2018) estimam uma taxa de erro de 3,5% para um total de 600 amostras aleatórias.

Figura 4 – (a) Etapas do processo executado pelo KT-Speech-Crawler e (b) Interface web para análise das amostras e suas transcrições



Fonte: adaptada de Lakomkin et al. (2018).

Na demonstração de resultados, Lakomkin *et al.* (2018) provaram que adicionando amostras capturadas do Youtube aumentaram positivamente a performance do reconhecimento de voz dos dois conjuntos de dados, Wall Street Journal e TED-LIUM v2. Foram adicionadas 200 horas de amostras do Youtube (108.617 dicções) no conjunto de treinamento do Wall Street Journal e a taxa de erro de palavras foi melhorada de 34,2% para 15,8%. Um resultado similar foi visto no conjunto de dados TED-LIUM v2 ao adicionar 300 horas de amostras do Youtube, melhorando a taxa de erro de caracteres de 10,4% para 8,2% (LAKOMKIN *et al.*, 2018).

### 3 PROPOSTA DO PROTÓTIPO

Nesta seção serão apresentadas as justificativas para a realização do trabalho proposto (subseção 3.1), bem como os requisitos principais (subseção 3.2) e finalizando com a metodologia e o cronograma planejado para o desenvolvimento do trabalho (subseção 3.3).

**Comentado [AS8]:** É a continuação da frase anterior? Ficou confuso.

### 3.1 JUSTIFICATIVA

Nas seções 1 e 2 foram evidenciadas a relevância do tema proposto. No Quadro 1 é apresentado um comparativo entre os trabalhos correlatos, de modo que as linhas representam as características e as colunas os trabalhos relacionados.

Quadro 1 - Comparativo entre os correlatos

Características	Correlatos	Uko, Dubukumah e Ayosubomi (2019)	Bailon <i>et al.</i> (2019)	Lakomkin <i>et al.</i> (2018)
Recuperar informações de vídeos do Youtube		X	X	✓
Reconhecimento de voz		✓	✓	✓
Transcrição de voz		✓	✓	✓
Tradução de texto		X	✓	X
Plataforma		Componente	Desktop	Desktop

Fonte: elaborado pelo autor.

Conforme demonstrado no Quadro 1, percebe-se que Uko, Dubukumah e Ayosubomi (2019), Bailon *et al.* (2019) e Lakomkin *et al.* (2018) têm finalidades semelhantes no que se diz respeito a reconhecimento e transcrição de voz. Ambas as características foram identificadas nos três correlatos como essenciais para alcançar seus objetivos, uma vez que são importantes para a operacionalização das demais características. Lakomkin *et al.* (2018) se destacam por apresentarem a característica relacionada a recuperar informações de vídeos do Youtube, dizendo respeito a recuperar dados da voz e outras informações.

O reconhecimento de voz é uma característica que foi vista em todos os trabalhos correlatos. Em Uko, Dubukumah e Ayosubomi (2019) a voz era recuperada por um dispositivo Android e transmitida pelo módulo Bluetooth para o componente desenvolvido. Bailon *et al.* (2019), para treinar a RNA, utilizavam-se do Windows Speech Recognition para a fim de reconhecer e melhorar o reconhecimento de voz por parte dos participantes. Já Lakomkin *et al.* (2018) fizeram uso das informações de vídeos do Youtube para obter a voz e treinar redes neurais artificiais.

Assim como a característica de reconhecimento de voz, a transcrição de voz também é fundamental nos três correlatos. Uko, Dubukumah e Ayosubomi (2019) precisam dessa característica para que o usuário saiba o que foi dito por outra pessoa. Bailon *et al.* (2019) necessitam transcrever a voz filipina para que seja possível traduzir o texto para chinês e manter a comunicação desses dois idiomas. Lakomkin *et al.* (2018) fizeram uso dessa característica para treinar redes neurais e conferir os resultados obtidos das amostras.

A característica de tradução de texto foi vista somente em Bailon *et al.* (2019). Ela se assemelha a este protótipo porque a pessoas que não falam o mesmo idioma possam entender o conteúdo da transmissão ao vivo. Bailon *et al.* (2019) têm seu principal objetivo voltado

**Comentado [AS9]:** O que é esta plataforma?

**Comentado [AS10]:** Colocar no início da frase.

**Comentado [AS11]:** Sempre colocar "Redes Neurais Artificiais", pois redes neurais são as que estão no nosso cérebro.

**Formatado:** Realce

para essa característica, na qual a comunicação entre filipino e chinês possa ser compreendida com resultados satisfatórios.

Com base nestas características e no Quadro 1 apresentado, nota-se que o trabalho proposto se torna relevante ao buscar ajudar pessoas a entenderem transmissões ao vivo que não são de seu idioma. Além disso, a proposta visa auxiliar em outros aspectos como a integração de novos conteúdos na vida das pessoas para que possam aprofundar estudos e até mesmo se divertir ao entenderem o que está acontecendo em transmissões ao vivo. Já como contribuição acadêmica será utilizado o Método RURUCAg, que poderá ser utilizado em trabalhos que queiram modelar a relação entre os requisitos da aplicação e práticas consolidadas no design de interface como as heurísticas de Nielsen, bem como avaliar a usabilidade e a experiência de usuários em aplicações na área da computação. Além disso, traz a contribuição acadêmica buscando encontrar interfaces amigáveis para serem aplicadas no protótipo, permitindo que os usuários finais se sintam mais confortáveis e-a ponto de? focarem nos seus objetivos. Como contribuição tecnológica pode-se destacar o desenvolvimento de uma aplicação web e aplicação móvel, com a tecnologia Progressive Web Application (PWA), e fazer uso de um Sistema Gerenciador de Banco de Dados (SGBD) Not Only Structured Query Language (NoSQL).

**Comentado [AS12]:** Esse texto está idêntico ao de outras propostas. Para isso deve ter citação direta ou indireta.

**Comentado [AS13]:** Usar essas tecnologias não são contribuições tecnológicas.

### 3.2 REQUISITOS PRINCIPAIS DO PROBLEMA A SER TRABALHADO

Nesta subseção serão abordados os principais Requisitos Funcionais (RF), assim como os principais Requisitos Não Funcionais (RNF), conforme Quadro 2.

Quadro 2 – Principais Requisitos Funcionais e Não Funcionais

A aplicação deverá:	Tipo
permitir ao usuário acesso por meio de uma conta Google.	RF
permitir ao usuário consultar uma lista de vídeos ao vivo do Youtube conforme as inscrições do usuário autenticado com a conta Google.	RF
permitir ao usuário visualizar um vídeo ao vivo.	RF
permitir ao usuário configurar a transcrição para o vídeo ao vivo sendo visualizado.	RF
utilizar serviços Google Cloud para transcrever áudio e traduzir essa transcrição.	RNF
utilizar APIs para recuperar informações de vídeos ao vivo do Youtube.	RNF
implementar a aplicação web utilizando o <i>framework</i> Angular.	RNF
implementar a aplicação web seguindo as diretrizes do Material Design (MD).	RNF
consultar serviços na parte do servidor desenvolvido em NodeJS.	RNF
implementar uma aplicação móvel com a tecnologia PWA.	RNF
implementar uma aplicação móvel com a ajuda do ambiente de desenvolvimento Visual Studio Code.	RNF
utilizar a plataforma Firebase (Google) para usar serviços de hospedagem e validar acessos de contas do Google.	RNF
utilizar a plataforma Firebase como SGBD NoSQL para armazenar informações dos usuários.	RNF
utilizar o Método RURUCAg para modelar os requisitos da aplicação com as heurísticas de Nielsen.	RNF
utilizar o Método RURUCAg para avaliar a usabilidade e a experiência de uso da aplicação.	RNF

Fonte: elaborado pelo autor.

**Comentado [AS14]:** Remover “.” Das frases.

### 3.3 METODOLOGIA

A metodologia desta proposta está elaborada em seis etapas e composta pelos instrumentos metodológicos contidos no Quadro 3, juntamente com os períodos relacionados.

**Quadro 3 – Cronograma com as etapas e metodologias**

Etapas	Quinzenas		2021									
			fev.		mar.		abr.		maio		jun.	
	1	2	1	2	1	2	1	2	1	2	1	2
<b>Pesquisa na literatura:</b> realizar uma revisão mais aprofundada da literatura sobre os assuntos citados na revisão bibliográfica e trabalhos correlatos.												
<b>Refinamento dos requisitos:</b> reavaliar os requisitos funcionais e não funcionais já definidos e, se necessário.												
<b>Especificação e análise:</b> formalizar as funcionalidades da ferramenta por meio de casos de uso e diagramas de atividade da Unified Modeling Language (UML), utilizando a ferramenta Astah Community.												
<b>Implementação da aplicação web:</b> implementar uma aplicação web para os usuários finais utilizando como base a linguagem de programação JavaScript e Visual Studio Code como ambiente de desenvolvimento.												
<b>Implementação da aplicação móvel:</b> implementar uma aplicação móvel com a tecnologia PWA para exibir o conteúdo da aplicação web e Visual Studio Code como ambiente de desenvolvimento.												
<b>Verificação e validação:</b> validar a usabilidade da solução pelo Método RURUCAg.												

Fonte: elaborado pelo autor.

**Comentado [AS15]:** A metodologia deve ser descrita fora do quadro. O quadro deve apresentar somente as etapas.

**Comentado [AS16]:** Início na segunda quinzena de fevereiro

## 4 REVISÃO BIBLIOGRÁFICA

Conforme os dados da Pesquisa Nacional por Amostra de Domicílios Contínua (PNAD) realizada no ano de 2016, uma das atividades principais do brasileiro na internet é assistir vídeos on-line no celular (IBGE, 2018). Segundo Nagumo, Teles e Silva (2020), a pesquisa Video Viewers realizada com três mil pessoas pelo Instituto Provokers, em 2018, mostrou que nove em cada dez brasileiros usam o Youtube para aprendizado. Com base no crescimento de acessos ao Youtube pelo mundo foi criada a Youtube Data API para simplificar e unificar a forma de capturar e enviar informações relacionadas a vídeos e canais no Youtube (GOOGLE, 2020).

O Google Tradutor é possivelmente o serviço de tradução mais utilizado, e em 2016, mais de 500 milhões de pessoas estava traduzindo 100 bilhões de palavras por dia em 103 idiomas, sendo o Brasil com maior número de usuários (BROOKS, 2016). Existem diferentes tecnologias que podem alcançar a implementação de fala e uma delas é a de Sistemas de Redes Neurais Artificiais (SRNA?) (BAILON *et al.*, 2019). Uma Rede Neural Recorrente (RNR) é uma arquitetura de RNA que contém conexões cíclicas, que permitem aprender a dinâmica temporal de dados sequenciais (MURAD; PYUN, 2017).

Para se ter uma maior interação e feedback com o usuário o uso de interfaces amigáveis é indispensável em uma aplicação desenvolvida, independente da tecnologia



utilizada (COSTA, 2018). Segundo Google (2019), o Material Design (MD) foi criado para que o usuário tenha menos esforço ao utilizar uma aplicação e consiga focar no que ele precisa. Torres *et al.* (2018) citam que o MD foi gerado para manter uma identidade visual em diferentes aplicações como em sites e no sistema operacional Android. De acordo com Costa (2018), o método RURUCAg foi criado para avaliar a interface de aplicações com base nas heurísticas de Nielsen e assim reconhecer incidentes com relação a usabilidade, visando assim melhorar a interface e funcionalidades da aplicação para o usuário final.

Por fim, cabe destacar que o aumento de projetos para aplicações móveis também surgiu a dificuldade de reutilização de código para aplicações existentes. O PWA foi criado para, de forma híbrida, unificar e minimizar implementações entre aplicações web e móveis. Para Sharma *et al.* (2019), utilizar o PWA implica em um baixo tempo de desenvolvimento e de implantação da aplicação, além de permitir que a empresa não contrate funcionários especializados em diferentes plataformas. As funcionalidades do PWA incluem: funcionamento off-line; sincronização em segundo plano; e instalação da aplicação web em plataformas móveis. Além de não precisar de uma loja de aplicativos para que a aplicação seja distribuída (BIØRN-HANSEN *et al.*, 2017).

#### REFERÊNCIAS

BAILON, Mark R. M.; DE SILVA, Mark A. C.; LAPUZ, Reyann J. P.; TINIO, John L. A.; YU, Ted B. K.; GUSTILO, Reggie C. **Filipino to chinese speech-to-speech translator using neural network with database system**. International Journal of Emerging Trends in Engineering Research, [S. l.], v. 7, n. 9, p. 276–282, 2019.

BARROZO, Ligia V.; SERAFIM, Mirela B.; MORAES, Sara L. de; MANSUR, Giselle. **MONITORAMENTO ESPAÇO-TEMPORAL DAS ÁREAS DE ALTO RISCO DE COVID-19 NOS MUNICÍPIOS DO BRASIL**. Hygeia - Revista Brasileira de Geografia Médica e da Saúde, [S. l.], p. 417–425, 2020.

BIØRN-HANSEN, Andreas; MAJCHRZAK, Tim A.; GRØNLI, Tor M. **Progressive web apps: The possible web-native unifier for mobile development**. In: WEBIST 2017 - PROCEEDINGS OF THE 13TH INTERNATIONAL CONFERENCE ON WEB INFORMATION SYSTEMS AND TECHNOLOGIES 2017, Anais [...]. [s.l.: s.n.].

BROOKS, Richard. **11 Google Translate Facts You Should Know**. K International, 02 mai. 2016. Disponível em: <https://k-international.com/blog/google-translate-facts/>. Acesso em: 14 out. 2020.

COSTA, Simone Erbs da. **iLibras como facilitador na comunicação do surdo: Desenvolvimento de um recurso colaborativo de tecnologia assistiva**. 2018. 263 f. Dissertação (Mestrado) – Centro de Ciências Tecnológicas, Universidade do Estado de Santa Catarina, 2018.

**Comentado [AS17]:** Falta um link entre as frases.

**Comentado [AS18]:** Confuso. Rever.

**Comentado [AS19]:** Seria a continuação da frase anterior?

**Comentado [AS20]:** Para 4 ou mais autores usar et al.

**Comentado [AS21]:** Não se coloca título em caixa alta.

**Comentado [AS22]:** Não usar caixa alta.

DESLANDES, Suely F.; COUTINHO, Tiago. **The intensive use of the internet by children and adolescents in the context of COVID-19 and the risks for self-inflicted violence.** *Ciência e Saúde Coletiva*, [S. l.], v. 25, p. 2479–2486, 2020.

GOOGLE. Material Design. **Introduction**. [S.l.], 2019. Disponível em: <https://material.io/design/introduction/>. Acesso em: 08 out. 2020.

GOOGLE. Youtube Data API. **Página inicial**, 2020. Disponível em: <https://developers.google.com/youtube/v3>. Acesso em: 08 out. 2020.

HAMID, Saima; MIR, Mohammad Y.; ROHELA, Gulab K. **Novel coronavirus disease (COVID-19): a pandemic (epidemiology, pathogenesis and potential therapeutics).** *New Microbes and New Infections*, 2020.

IBGE. **Acesso à internet e à televisão e posse de telefone móvel celular para uso pessoal: 2016.** Rio de Janeiro: IBGE, 2018.

LAKOMKIN, Egor; MAGG, Sven; WEBER, Cornelius; WERMTER, Stefan. **KT-Speech-Crawler: Automatic dataset construction for speech recognition from YouTube videos.** *EMNLP 2018 - CONFERENCE ON EMPIRICAL METHODS IN NATURAL LANGUAGE PROCESSING: SYSTEM DEMONSTRATIONS, PROCEEDINGS* 2018.

MURAD, Abdulmajid; PYUN, Jae Y. **Deep recurrent neural networks for human activity recognition.** *Sensors (Switzerland)*, [S. l.], v. 17, n. 11, 2017.

NAGUMO, Estevon; TELES, Lúcio F.; SILVA, Lucélia de A. **A utilização de vídeos do Youtube como suporte ao processo de aprendizagem (Using Youtube videos to support the learning process).** *Revista Eletrônica de Educação*, [S. l.], v. 14, p. 3757008, 2020.

SHARMA, Varsha; VERMA, Rajat; PATHAK, Vaishali; PALIWAL, Muskan; JAIN, Priya. **Progressive Web App (PWA) - One Stop Solution for All Application Development Across All Platforms.** *International Journal of Scientific Research in Computer Science, Engineering and Information Technology*, [S. l.], p. 1120–1122, 2019.

TORRES, Cecília V. et al. **DESENVOLVIMENTO MOBILE COM ENFOQUE ACESSÍVEL: O DESIGN NA MEDIAÇÃO DA INCLUSÃO.** *Human Factors in Design*, [S. l.], v. 7, n. 13, p. 85–101, 2018.

UKO, Victor S.; DUBUKUMAH, Gachada B.; AYOSUBOMI, Ibrahim K. **Microcontroller Based Speech to Text Translation System.** *European Journal of Engineering Research and Science*, [S. l.], v. 4, n. 12, p. 149–154, 2019.

XAVIER, Marcelle B.; CARVALHO, Francisco dos S.; CARVALHO, Mauro dos S.; MORAES, Juliana M. de. **Identidade Surda: Uma Análise dos Discursos de Professores Surdos e Ouvintes no Youtube / Deaf Identity: An Analysis of the Discourses of Deaf Teachers and Listeners on Youtube.** *ID on line REVISTA DE PSICOLOGIA*, [S. l.], v. 13, n. 45, 2019.

**Comentado [AS23]:** Verificar a norma para referenciar o mesmo autor mais de 1 vez. Corrigir no texto também.

**Comentado [AS24]:** Não usar caixa alta.

**Comentado [AS25]:** Para 4 ou mais autores usar et al.

**ASSINATURAS**

(Atenção: todas as folhas devem estar rubricadas)

Assinatura do(a) Aluno(a): \_\_\_\_\_

Assinatura do(a) Orientador(a): \_\_\_\_\_

Assinatura do(a) Coorientador(a) (se houver): \_\_\_\_\_

Observações do orientador em relação a itens não atendidos do pré-projeto (se houver):

## FORMULÁRIO DE AVALIAÇÃO – PROFESSOR TCC I

Acadêmico(a): Hugo Marcel Larsen \_\_\_\_\_

Avaliador(a): Andreza Sartori \_\_\_\_\_

ASPECTOS AVALIADOS <sup>1</sup>		atende	atende parcialmente	não atende
ASPECTOS TÉCNICOS	1. INTRODUÇÃO O tema de pesquisa está devidamente contextualizado/delimitado?	X		
	O problema está claramente formulado?	X		
	2. OBJETIVOS O objetivo principal está claramente definido e é passível de ser alcançado?	x		
	Os objetivos específicos são coerentes com o objetivo principal?		X	
	3. JUSTIFICATIVA São apresentados argumentos científicos, técnicos ou metodológicos que justificam a proposta?	X		
	São apresentadas as contribuições teóricas, práticas ou sociais que justificam a proposta?	X		
	4. METODOLOGIA Foram relacionadas todas as etapas necessárias para o desenvolvimento do TCC?	x		
ASPECTOS METODOLÓGICOS	Os métodos, recursos e o cronograma estão devidamente apresentados?		x	
	5. REVISÃO BIBLIOGRÁFICA (atenção para a diferença de conteúdo entre projeto e pré-projeto) Os assuntos apresentados são suficientes e têm relação com o tema do TCC?	x		
	6. LINGUAGEM USADA (redação) O texto completo é coerente e redigido corretamente em língua portuguesa, usando linguagem formal/científica?	X		
	A exposição do assunto é ordenada (as ideias estão bem encadeadas e a linguagem utilizada é clara)?	X		
	7. ORGANIZAÇÃO E APRESENTAÇÃO GRÁFICA DO TEXTO A organização e apresentação dos capítulos, seções, subseções e parágrafos estão de acordo com o modelo estabelecido?		x	
	8. ILUSTRAÇÕES (figuras, quadros, tabelas) As ilustrações são legíveis e obedecem às normas da ABNT?	X		
	9. REFERÊNCIAS E CITAÇÕES As referências obedecem às normas da ABNT?		x	
As citações obedecem às normas da ABNT?			x	
Todos os documentos citados foram referenciados e vice-versa, isto é, as citações e referências são consistentes?			x	

### PARECER – PROFESSOR DE TCC I OU COORDENADOR DE TCC (PREENCHER APENAS NO PROJETO):

O projeto de TCC será reprovado se:

- qualquer um dos itens tiver resposta NÃO ATENDE;
- pelo menos **4 (quatro)** itens dos **ASPECTOS TÉCNICOS** tiverem resposta ATENDE PARCIALMENTE; ou
- pelo menos **4 (quatro)** itens dos **ASPECTOS METODOLÓGICOS** tiverem resposta ATENDE PARCIALMENTE.

**PARECER:** ( ) APROVADO ( ) REPROVADO

Assinatura: \_\_\_\_\_ Data: 04/11/2020 \_\_\_\_\_

<sup>1</sup> Quando o avaliador marcar algum item como atende parcialmente ou não atende, deve obrigatoriamente indicar os motivos no texto, para que o aluno saiba o porquê da avaliação.