TELL YOUR WORDS

SPEECH TRAINER FOR CHILDREN

KIM JOONHEE, DEPT OF INFORMATION SYSTEM 18
MA YUDAM, DEPT OF KOREAN LANGUAGE & LITERATURE 19
PARK JUNHYEONG, DEPT OF BUSINESS ADMINISTRATION 18

CONTENTS

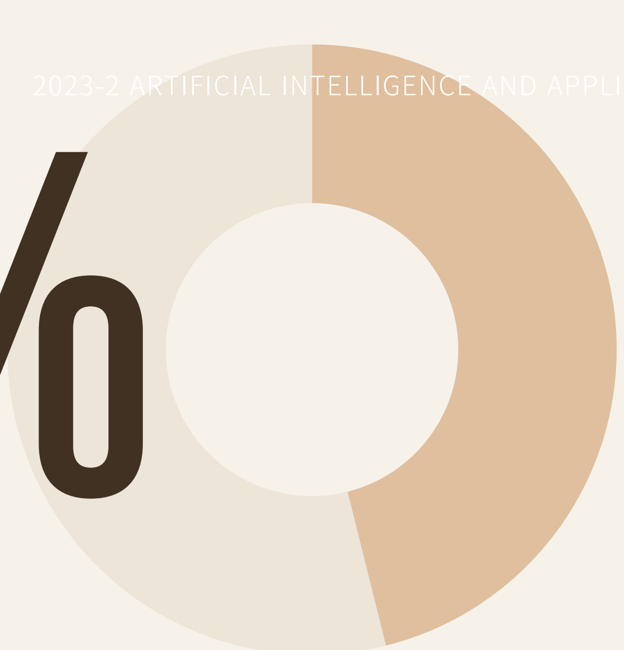
SERVICE INTRODUCTION

WHAT KIND OF PROBLEM WE FACE?

46.1% OF DOUBLE-INCOME HOUSEHOLDS... HIGHEST EVER

THE PROPORTION OF DUAL-INCOME HOUSEHOLDS IS THE HIGHEST SINCE 2015 WHEN RELATED STATISTICS BEGAN TO BE COMPILED

46.1%

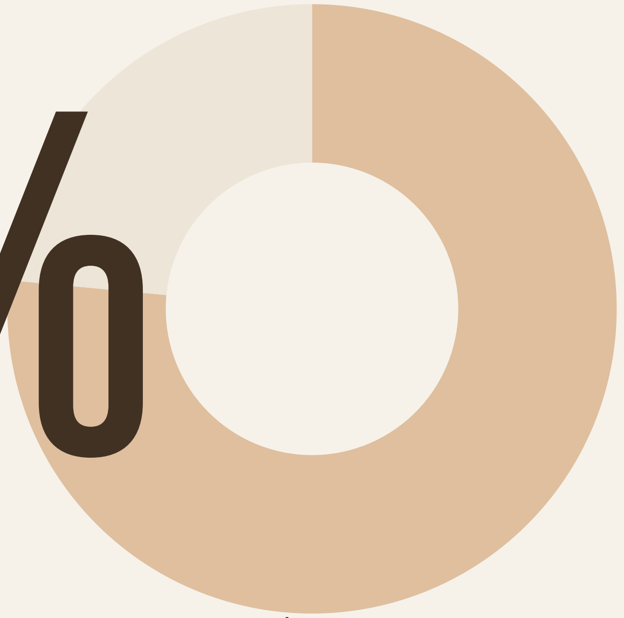


SOURCE) NATIONAL STATISTICAL OFFICE OF KOREA

IN THE AFTERMATH OF COVID-19, 'I EXPERIENCED A CHILD-REARING GAP'

IN PARTICULAR, DUAL-INCOME OFFICE WORKERS WITH INFANTS (AGES 4 TO 7) WERE THE HIGHEST AT 90.4%

76.5%

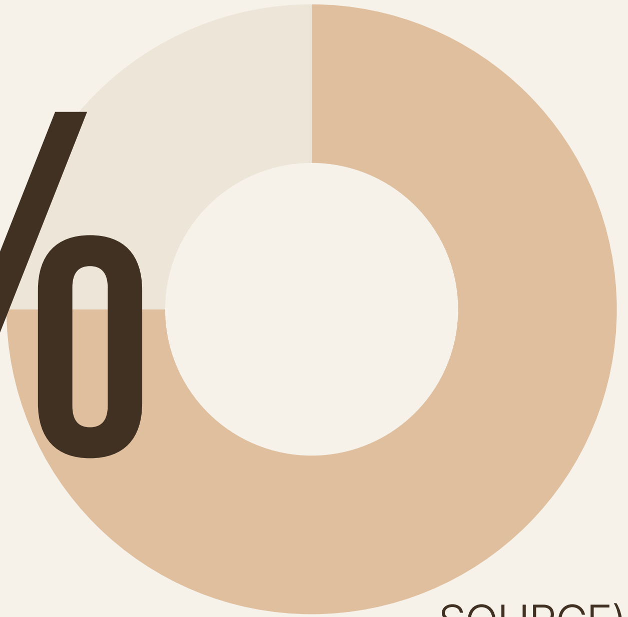


SOURCE) INCRUIT X ALBACALL

'CHILDREN'S CHANCES OF DEVELOPING LANGUAGE SKILLS HAVE DECREASED'

75% OF OFFICIALS AT NATIONAL AND PUBLIC DAYCARE CENTERS RESPONDED, AND LANGUAGE DISORDERS INCREASED 30% COMPARED TO BEFORE COVID-19.

75%



SOURCE) EBS

WHAT KIND OF PROBLEM WE FACE?

COVID-19

DUE TO THE SOCIAL DISTANCE, CHILDREN HAVE A PROBLEM OF LACK OF LANGUAGE SKILLS

DOUBLE-INCOME FAMILIES

FOR THESE CHILDREN, ADULTS HAVE TO SUPPORT THEM BUT IT'S HARD BECAUSE DOUBLE-INCOME FAMILIES ARE COMMON, ESPECIALLY IN KOREA

THE DIRECTION OF PRIVATE EDUCATION

ACADEMIES ONLY DO FOCUS ON GRAMMAR AND UNIVERSITY ENROLLMENT EXAMS RATHER THAN CHILDREN'S PRONUNCIATION AND SPEAKING SKILLS

WHAT'S NOT IN THE KOREAN ENGLISH EDUCATION MARKET

FURTHERMORE, THIS IDEA MAY TARGET THE RELATIVE LACK OF SPEAKING EDUCATION IN THE ENGLISH EDUCATION MARKET IN KOREA

WE SUGGEST A SPEAKING PRACTICE SERVICE FOR CHILDREN

A

NUGU SAYS,
“REPEAT AFTER ME“

B

CHILD REPEAT
THE SENTENCES

C

CHECK
PRONUNCIATION

D

SEND SCORE
AS A REPORT
TO THEIR PARENTS

WHICH SENTENCES DO WE TEACH?

LEARNING SOUNDS CAN BE DIFFERENT DEPENDING ON
HOW WE STUDY AND WHAT SOUNDS WE'RE LOOKING AT

FOR EXAMPLE, THE SOUND /L/ IS LEARNED LATER
USUALLY, LIQUIDS GET REPLACED BY GLIDES,
IN ENGLISH WHERE /R/ BECOMES [W] AND /L/ BECOMES [W] OR [Y]

FREQUENT PRONUNCIATION ERROR EXAMPLES IN ENGLISH

[WÆBɪT] 'RABBIT' ([R] -> [W])
[FIT] ‘PIT’([P] -> [F])
[YIF] 'LEAF'([L] -> [Y])

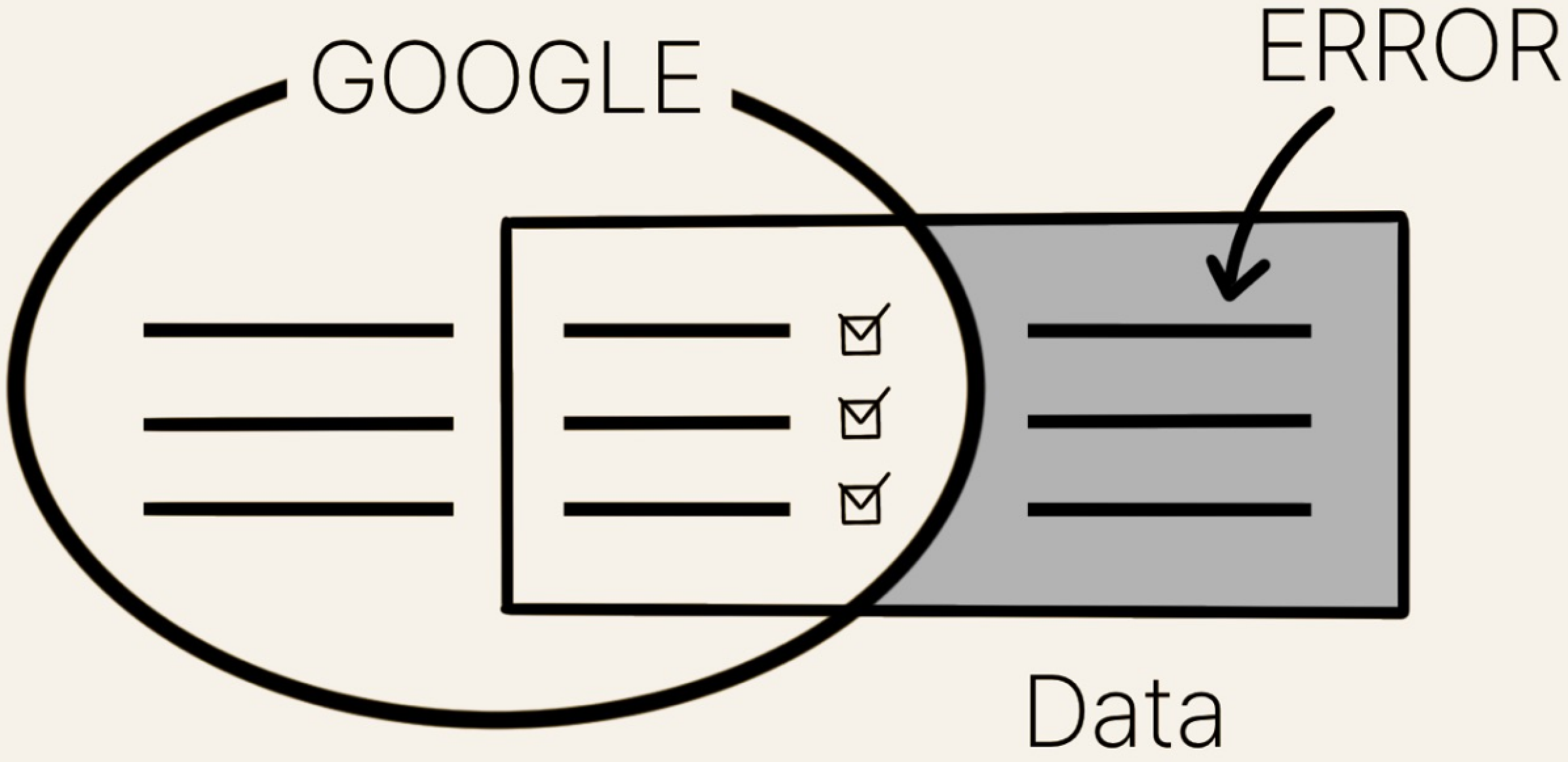
WHICH SENTENCES DO WE TEACH?

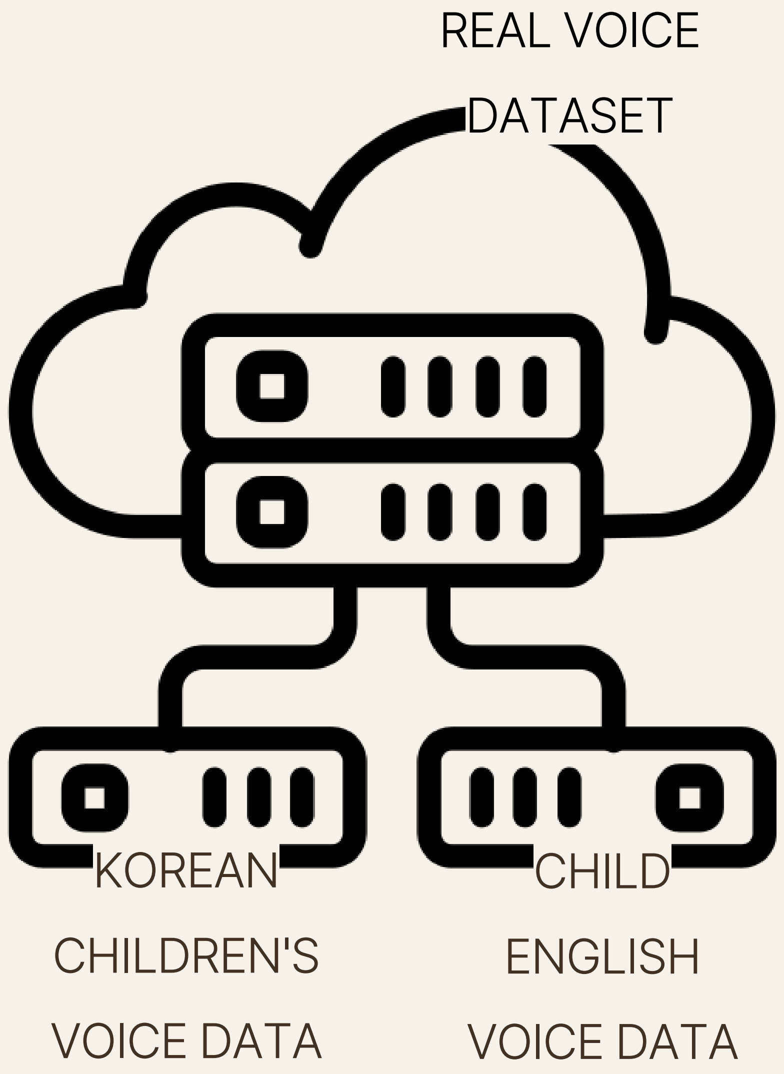
IN KOREAN, ALSO SIMILAR MISTAKES OCCUR.
BUT IN ‘REPLACING’, IT DEPENDS ON LANGUAGE STRUCTURE.

FREQUENT PRONUNCIATION ERROR EXAMPLES IN KOREAN

[POWICHA] /POLICHA/ ‘보리차’ ([리] -> [위])
[YA:MƆN] /LAMYƏN/ ‘라면’ ([라] -> [야])
[THAYAM] /SALAM/ ‘사람’ ([사] -> [따])
[KITA] /KICHA/ ‘기차' ([차] -> [타])

GETTING SENTENCES THAT ARE WORTH REPEATING TO CHILDREN





GETTING SENTENCES THAT ARE WORTH REPEATING TO CHILDREN

```
recognizer = sr.Recognizer()
recognized_texts = []

# Load wav files
for wav_file in glob.glob(wav_files_path):
    try:
        with sr.AudioFile(wav_file) as source:
            audio_data = recognizer.record(source)
        # STT by Google Web Speech API
        text = recognizer.recognize_google(audio_data, language="default")
        # language = 'ko-KR' when you want to use for korean sentences

        recognized_texts.append({'filename': wav_file, 'text': text})
    except sr.RequestError as e:
        print(f"Google Web Speech API request error: {e}")
    except Exception as e:
        print(f"Error processing {wav_file}: {e}")

# Create a DataFrame from the list of recognized texts
recognized_texts_df = pd.DataFrame(recognized_texts)
```

SINCE 'CHILDREN'S VOICE DATA' HAS BEEN CALIBRATED AND LABELED BY HUMANS,

THE TRANSCRIPT DATA MADE FROM THIS DATA IS VIEWED AS 'CORRECTED_TEXT'

AND COMPARED WITH THE SENTENCES RECEIVED BY THE STT MODEL.

GETTING SENTENCES THAT ARE WORTH REPEATING TO CHILDREN

```
json_texts_df = pd.DataFrame()

# Load json files
for json_file in glob.glob(json_files_path):
    with open(json_file, 'r') as file:
        data = pd.json_normalize(json.load(file))
        json_texts_df = json_texts_df.append(data, ignore_index=True)

json_texts_df.head()
corrected_texts_df = json_texts_df[['File.FileName', 'Transcription.LabelText']]
corrected_texts_df = corrected_texts_df.rename(columns={'File.FileName': 'filename', 'Transcription

merged_df = pd.merge(corrected_texts_df, recognized_texts_df, on=['filename'], how='inner')
```

```
merged_df['corrected_text_cleaned'] = merged_df['corrected_text'].str.replace(r'[\s\W_]+', '')
merged_df['text_cleaned'] = merged_df['text'].str.replace(r'[\s\W_]+', '')
```

LOADING JSON FILES AND MAKE DATAFRAME WITH ONLY THE SENTENCES WITH THE ERROR CORRECTED IN THE JSON FORMAT FILE

IT IS NECESSARY TO PREPROCESS DATA FRAMES MADE USING GOOGLE API AND JSON FILES SO THAT THEY CAN BE PROCESSED

THESE SENTENCES COULD BE REASONABLE ONES

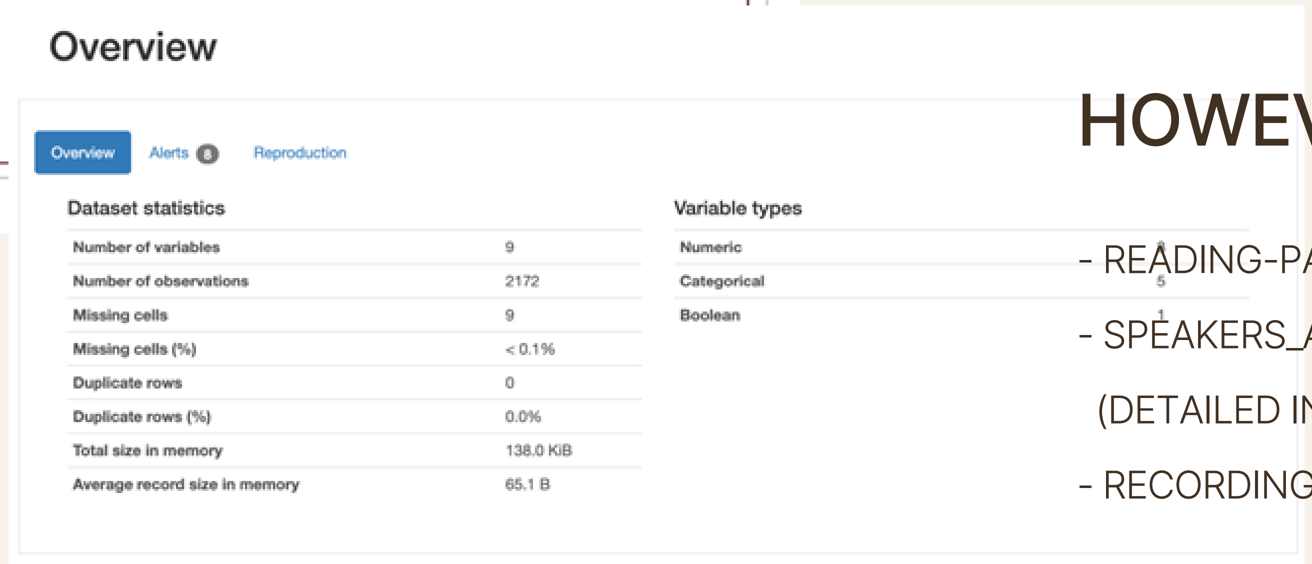
ENGLISH SENTENCE EXAMPLES

A. THE HAPPY CAT SAT ON THE MAT
B. HE HAS A BIG BLUE BALL TO PLAY
C. THE LITTLE DUCK QUACKS IN THE POND
D. THE SUN SHINES IN THE SKY
E. TIM'S TOY TRUCK IS GREEN

ENGLISH SENTENCE EXAMPLES

A. 오리도 만들고 기차도 만들었다
B. 아침에 라면을 먹어 행복했다
C. 공연하는 사람들은 다 키가 엄청나게 컸다
D. 코끼리 미끄럼틀을 탔다
E. 구름 사이로 우리 비행기가 슝 지나갔다

DATA DESCRIPTION


Overview

Overview | Alerts 8 | Reproduction

Dataset statistics

Number of variables	9
Number of observations	2172
Missing cells	9
Missing cells (%)	< 0.1%
Duplicate rows	0
Duplicate rows (%)	0.0%
Total size in memory	138.0 KiB
Average record size in memory	65.1 B

Variable types

Numeric
Categorical
Boolean

ENGLISH SPEECH DATA RECORDED BY PEOPLE PEOPLE OF DIFFERENT COUNTRIES, GENDERS, AND AGES. CONTAINING **2,140** SPEECH SAMPLES

RESEARCHERS CONSTRUCTED AN ELICITATION PARAGRAPH AND EACH SUBJECT'S PRONUNCIATION IS RECORDED

SINCE THE SAME PARAGRAPH WAS RECORDED BY MULTIPLE PEOPLE WITH DIFFERENT ENGLISH LEARNING BACKGROUNDS, THIS DATASET IS FIT TO OUR TASK

HOWEVER, IT LACKS PRONUNCIATION SCORE

- READING-PASSAGE.TXT: AN ELICITATIOIN PARAGRAPH
- SPEAKERS_ALL.CSV: DEMOGRAPHIC INFORMATION ON EVERY SPEAKER (DETAILED INFORMATION IN SPEECH_DATASET_PROFILE.HTML)
- RECORDING: A ZIPPED FOLDER CONTAINING .MP3 FILES WITH SPEECH

SOURCE: WEINBERGER, STEVEN., 2015. SPEECH ACCENT ARCHIVE. GEORGE MASON UNIVERSITY. RETRIEVED FROM HTTP://ACCENT.GMU.EDU

LABELING
VIA FEW SHOT LEARNING

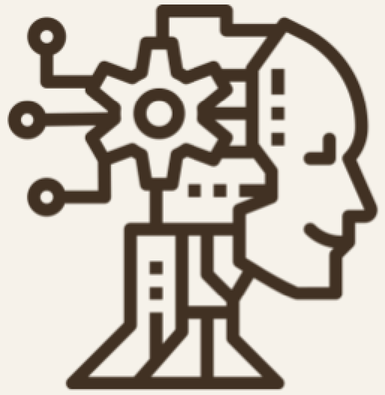
SINCE THE DATASET LACKS THE PRONUNCIATION SCORE,
WE DECIDED TO UTILIZE ***FEW SHOT LEARNING** TECHNIQUE
TO LABEL THE PRONUNCIATION SCORE OF EACH SPEECH DATA

*FEW SHOT LEARNING: A MACHINE LEARNING PARADIGM
WHERE A MODEL IS TRAINED TO MAKE ACCURATE PREDICTIONS
WITH ONLY A SMALL NUMBER OF EXAMPLES PER CLASS

1) MANUAL LABELING

MANUALLY LABELED
THE PRONUNCIATION SCORE
OF SAMPLE DATA AS 0, 1, 2

2) TRAIN FEW SHOT MODEL



USING THE MANUALLY
LABELED DATA,
TRAIN THE FEW SHOT MODEL

3) LABELING VIA FEW SHOT MODEL

TRAINED FEW SHOT MODEL
PREDICTS PRONUNCIATION SCORE
OF EACH SPEECH DATA

LABELING
VIA FEW SHOT LEARNING

1) MANUAL LABELING

MANUALLY LABELED SAMPLE SPEECH DATA AS 0(BAD), 1(NORMAL), 2(GOOD) FOR EACH OF FOUR METRICS

WE MANUALLY LABELED 10 SAMPLE DATA FOR EACH EVALUATION METRIC AND PRONUNCIATION LEVEL, TOTAL 120 SAMPLE DATA

ONLY LABELED DATA WITH CLEAR CLASSIFICATIONS

LABELING
VIA FEW SHOT LEARNING

1) MANUAL LABELING

FOUR REPRESENTATIVE PRONUNCIATION METRICS

Metrics	Description
Accuracy	The level of the learner pronounce each word in the utterance correctly
Completeness	The percentage of the words that are actually pronounced
Fluency	Does the speaker pronounce smoothly and without unnecessary pauses?
Prosodic	Does the speaker pronounce in correct intonation, stable speaking speed and rythm?

LABELING VIA FEW SHOT LEARNING

2) TRAIN FEW SHOT MODEL & 3) LABELING VIA FEW SHOT MODEL

USING THE MANUALLY LABELED DATA, WE TRAINED THE FEW SHOT MODEL

SET THE LABELED DATA AS TRAIN SET, AND NOT LABELED DATA SET AS TEST SET

FOR REAL WORLD SERVICES, IT WOULD BE USEFUL USE FEW-SHOT LEARNING BASED ON DATA LABELED BY EXPERTS

LABELING
VIA FEW SHOT LEARNING

2) TRAIN FEW SHOT MODEL &
3) LABELING VIA FEW SHOT MODEL

1

```
ss Audio_Encoder(nn.Module):
    def __init__(self, num_heads, num_layers):
        super().__init__()
        self.sentecne_level = nn.TransformerEncoder(nn.Transformer

    def forward(self, batch):
        max_len = max([e.size(0) for e in batch])
        padded_embeddings = torch.zeros(len(batch), max_len, batch
        for i, emb in enumerate(batch):
            seq_len = emb.size(0)
            padded_embeddings[i, :seq_len, :] = emb
        random_tensor = torch.randn(padded_embeddings.size(0), 1,
        batch_tensor = torch.cat((random_tensor, padded_embeddings
        batch_tensor = batch_tensor.permute(1 ,0 ,2).float()
        padding_mask = batch_tensor.sum(dim=-1).permute(1 ,0) == 0
        output_batch = self.sentecne_level(batch_tensor.float(), s
        output_batch = output_batch.permute(1 ,0 ,2)
        feature_vecs = output_batch[:,0,:]
        return feature_vecs
```

DEFINE A CUSTOMIZED

ENCODER CLASS THAT LEVERAGES

TRANSFORMER'S ENCODER TO

EXTRACT HIGH-DIMENSIONAL

FEATURE VECTORS

LABELING
VIA FEW SHOT LEARNING

2) TRAIN FEW SHOT MODEL &
3) LABELING VIA FEW SHOT MODEL

2

```
ass few_shot_Model(nn.Module):
    def __init__(self, encoder):
        super().__init__()
        self.encoder = encoder
        self.fc1 = nn.Linear(768,768)
        self.ac = nn.ReLU()
        self.fc2 = nn.Linear(768,256)
        self.sigmoid = nn.Sigmoid()

    def forward(self, voice_pair):
        voice_pair = self.encoder(voice_pair)
        voice_pair = self.fc1(voice_pair)
        voice_pair = self.ac(voice_pair)
        voice_pair = self.fc2(voice_pair)
        similarity = torch.cosine_similarity(voice_pair[0
        out = similarity
        return out
```

BUILDING A FEW SHOT MODEL

USING THE ENCODER CLASS

IT CONSISTS OF ONE ENCODER LAYER

AND TWO FULLY CONNECTED LAYERS

LABELING
VIA FEW SHOT LEARNING

2) TRAIN FEW SHOT MODEL &
3) LABELING VIA FEW SHOT MODEL

3

```
# data with same classes
for epoch in range(10) :
    epoch_loss = 0
    target = torch.tensor([1.0]).to(device)

    total_loss = 0
    optimizer.zero_grad()
    for i,j in combinations(g1,2) :
        f_model.train()
        data = df.iloc[[i,j],2]
        data = tuple(d.to(device) for d in data)
        output = f_model(data).unsqueeze(0)
        loss = criterion(output, target)
        total_loss += loss
    total_loss.backward()
    optimizer.step()
    epoch_loss += total_loss
```

TRAIN THE MODEL ON

DATA OF SAME CLASS

AND DATA OF DIFFERENT CLASS

LABELING
VIA FEW SHOT LEARNING

2) TRAIN FEW SHOT MODEL &
3) LABELING VIA FEW SHOT MODEL

4

```
for idx in tqdm(tests):
    f_model.eval()
    s1 = 0
    for i1 in g1 :
        data = df.iloc[[idx, i1], 2]
        data = tuple(d.to(device) for d in data)
        s1 += torch.sigmoid(f_model(data)).detach().i

    s2 = 0
    for i1 in g2 :
        data = df.iloc[[idx, i1], 2]
        data = tuple(d.to(device) for d in data)
        s2 += torch.sigmoid(f_model(data)).detach().i

    s3 = 0
    for i1 in g3 :
        data = df.iloc[[idx, i1], 2]
```

PREDICT THE PRONUNCIATION SCORE

OF EACH SPEECH DATASET

AUDIO AUGMENTATION

THE NUMBER OF OUR TOTAL SPEECH DATA WAS NEARLY 2,000,
WE NEEDED MORE DATA TO TRAIN OUR DEEP LEARNING MODEL

SINCE THE ORIGINAL DATASET WAS RECORDED IN A QUIET ROOM,
AUDIO AUGMENTATION CAN IMPROVE THE MODEL PERFORMANCE

TO AUGMENT AUDIO FILES, WE REFERRED TO “DATA AUGMENTING CONTRASTIVE LEARNING OF SPEECH REPRESENTATIONS IN THE TIME DOMAIN” (KHARITONOV ET AL., 2020) FROM PAPERSWITHCODE” AND USED WAVAUGMENT LIBRARY

EXECUTED THE AUGMENTATION ONCE FOR EACH DATA,
RESULTING IN A TOTAL OF NEARLY 4,000 DATASETS

AUDIO AUGMENTATION

Modification	Description
Pitch shift	Make lower or higher the pitch of the voice
Reverberation	Add echo to a sound signal, conveying spatial depth and width to the sound
Noise	Applying additive noise In this case, we used generated uniform noise
Time dropout	Substituting a brief segment of audio with periods of silence This method is frequently employed in the literature

APPROACH 1

WE TOOK TWO DIFFERENT APPROACHES TO PREDICT THE CHILDREN'S PRONUNCIATION
ONE BASED ON THE COSINE SIMILARITY COMPARISON,
AND THE OTHER BASED ON THE FINE-TUNED MODEL PREDICTION

A BIG ASSUMPTION IS THAT AUDIO FILES WITH SIMILAR PRONUNCIATION
WILL ALSO BE SIMILAR WHEN THEY ARE VECTORIZED

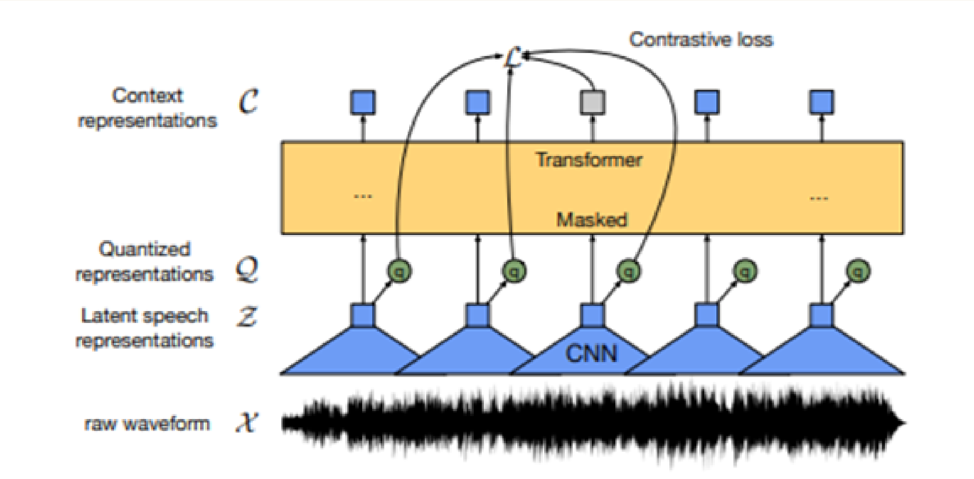
IF WE HAVE A REFERENCE DATA THAT CONSISTS OF VARIOUS PRONUNCIATIONS ON THE SAME PHRASE,
WE CAN DETERMINE THE SCORE OF NEW INPUT DATA BASED ON THE REFERENCE

FOR THE USE OF *WAV2VEC2 MODEL, WE REFERED “WAV2VEC 2.0: A FRAMEWORK FOR SELF-SUPERVISED
LEARNING OF SPEECH REPRESENTATIONS” (BAEVSKI ET AL., 2020) FROM PAPERSWITHCODE

*WAV2VEC 2.0: MODEL DEVELOPED BY FACEBOOK, THAT LEARNS POWERFUL REPRESENTATION FROM SPEECH AUDIO
IT IS PRE-TRAINED WITH 53,000 HOURS OF SPEECH DATA, AND IT CAN BE FINE-TUNED FOR DOWNSTREAM TASKS

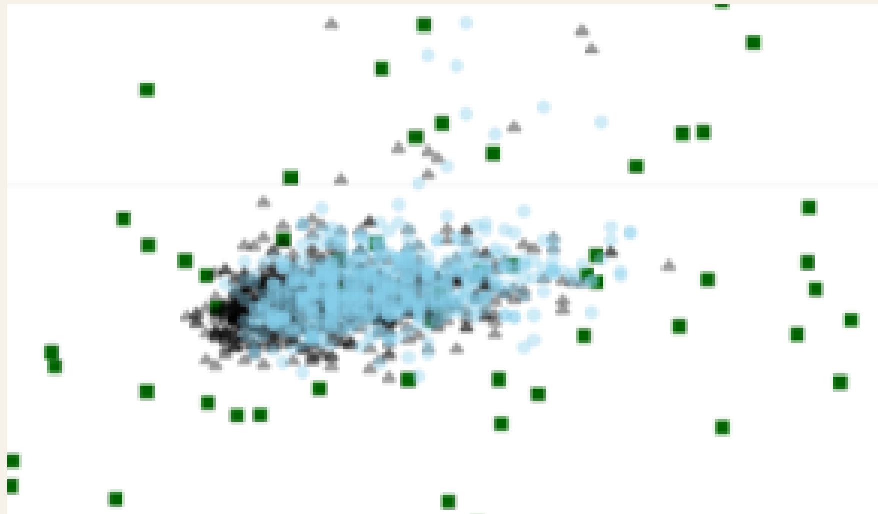
APPROACH 1

1



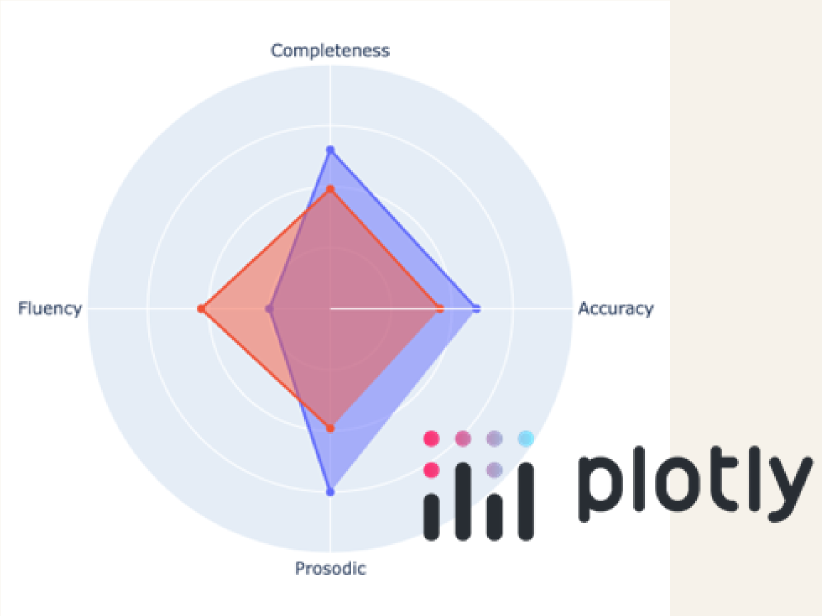
CONVERT THE RE
FERENCE AUDIO FILES TO
A TENSOR USING WAV2VEC 2.0 MODEL

2



CONVERT TEST DATA
INTO TENSOR AND FIND THE
N MOST SIMILAR TENSORS

3



VISUALIZE A CHILD'S
PREDICTED PRONUNCIATION SCORE
USING PLOTLY LIBRARY

APPROACH 2

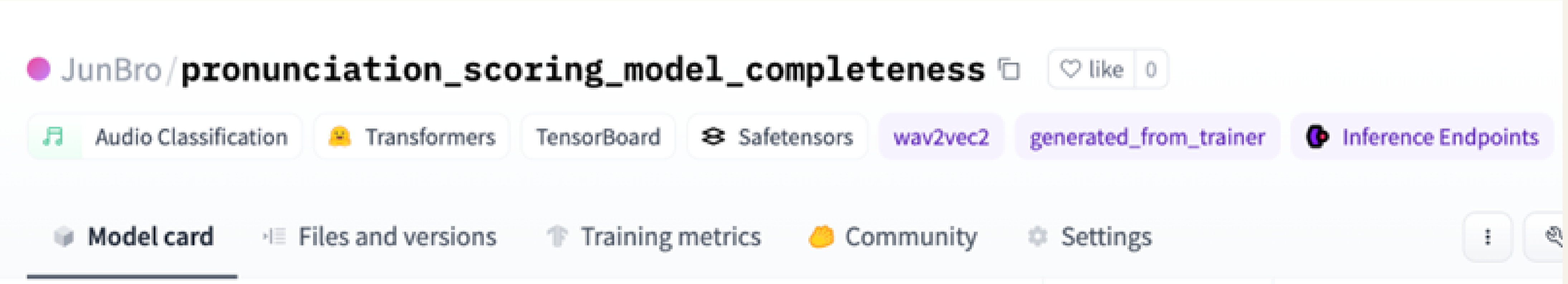
FINE-TUNED WAV2VEC 2.0 MODEL
FOR PRONUNCIATION SCORE PREDICTION

USED LABELED DATA FROM THE FEW SHOT LEARNING TO FINE TUNE THE WAV2VEC 2.0 MODEL
AND SPLIT THEM INTO TRAIN, VALID AND TEST SET

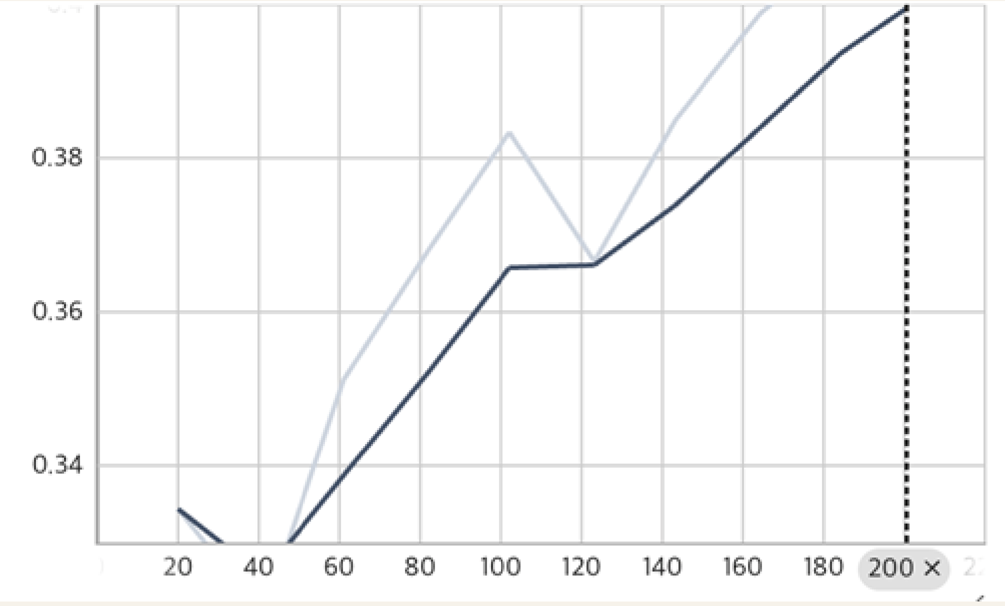
UTILIZED HUGGING FACE’S LIBRARIES AND UPLOADED THE TRAINED MODEL
TO OUR HUGGING FACE’S MODEL SPACE (REFER TO OUR TECH BLOG)

SINCE THERE ARE FOUR TARGET VARIABLES, THERE ARE FOUR DIFFERENT MODELS

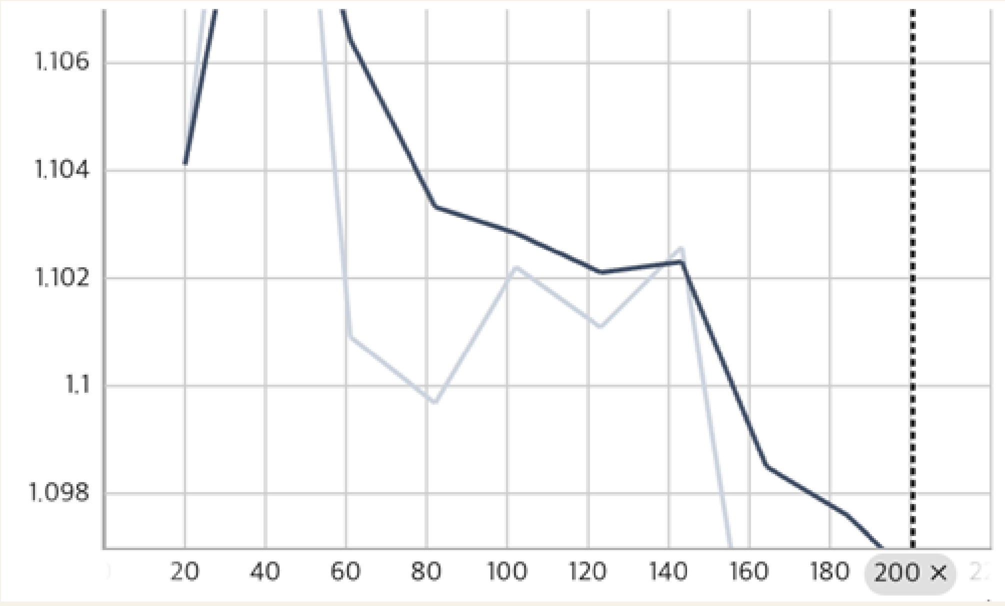
APPROACH 2



ACCURACY



LOSS



TRACKING THE TRAINING HISTORY

1) PREPROCESS THE DATASET IN SUITABLE FORMAT FOR MODEL TRAINING

2) FINE TUNING PRE-TRAINED MODEL IN FOUR DIFFERENT VERSIONS

3) USING FINE-TUNED MODELS, PREDICT THE PRONUNCIATION SCORE OF TEST FILE, AND VISUALIZE

plotly

EVALUATION

WE EVALUATED THE MODEL PERFORMANCE AND VISUALIZED THE RESULT

COMPARING THE PREDICTION AND THE ORIGINAL LABELS,
WE MADE CONFUSION MATRIX AND VISUALIZE IT AS A HEATMAP

TO CHECK VARIOUS CLASSIFICATION METRICS RESULTS,
WE USED CLASSIFICATION REPORT METHOD FROM SCIKIT-LEARN

MODEL TENDS TO BE GOOD AT CATCHING ‘BAD’ PRONUNCIATION,
SINCE IT IS VERY DISTINCTIVE FROM THE OTHERS

IF THE MANUAL LABELING STEP DONE THOROUGHLY BY EXPERTS,
AND IF THE MANUALLY LABELED DATASET IS LARGE ENOUGH,
THE MODEL CAN BE EXPECTED TO PERFORM BETTER

EVALUATION

	precision	recall	f1-score	support
Bad	0.21	0.27	0.24	26
Normal	0.00	0.00	0.00	31
Good	0.40	0.53	0.46	43
accuracy			0.30	100
macro avg	0.21	0.27	0.23	100
weighted avg	0.23	0.30	0.26	100

ACCURACY

	precision	recall	f1-score	support
Bad	0.38	0.45	0.41	33
Normal	0.35	0.53	0.42	36
Good	0.50	0.10	0.16	31
accuracy			0.37	100
macro avg	0.41	0.36	0.33	100
weighted avg	0.41	0.37	0.34	100

COMPLETENESS

	precision	recall	f1-score	support
Bad	0.41	0.63	0.49	35
Normal	0.40	0.55	0.46	29
Good	0.17	0.03	0.05	36
accuracy			0.39	100
macro avg	0.32	0.40	0.34	100
weighted avg	0.32	0.39	0.32	100

FLUENCY

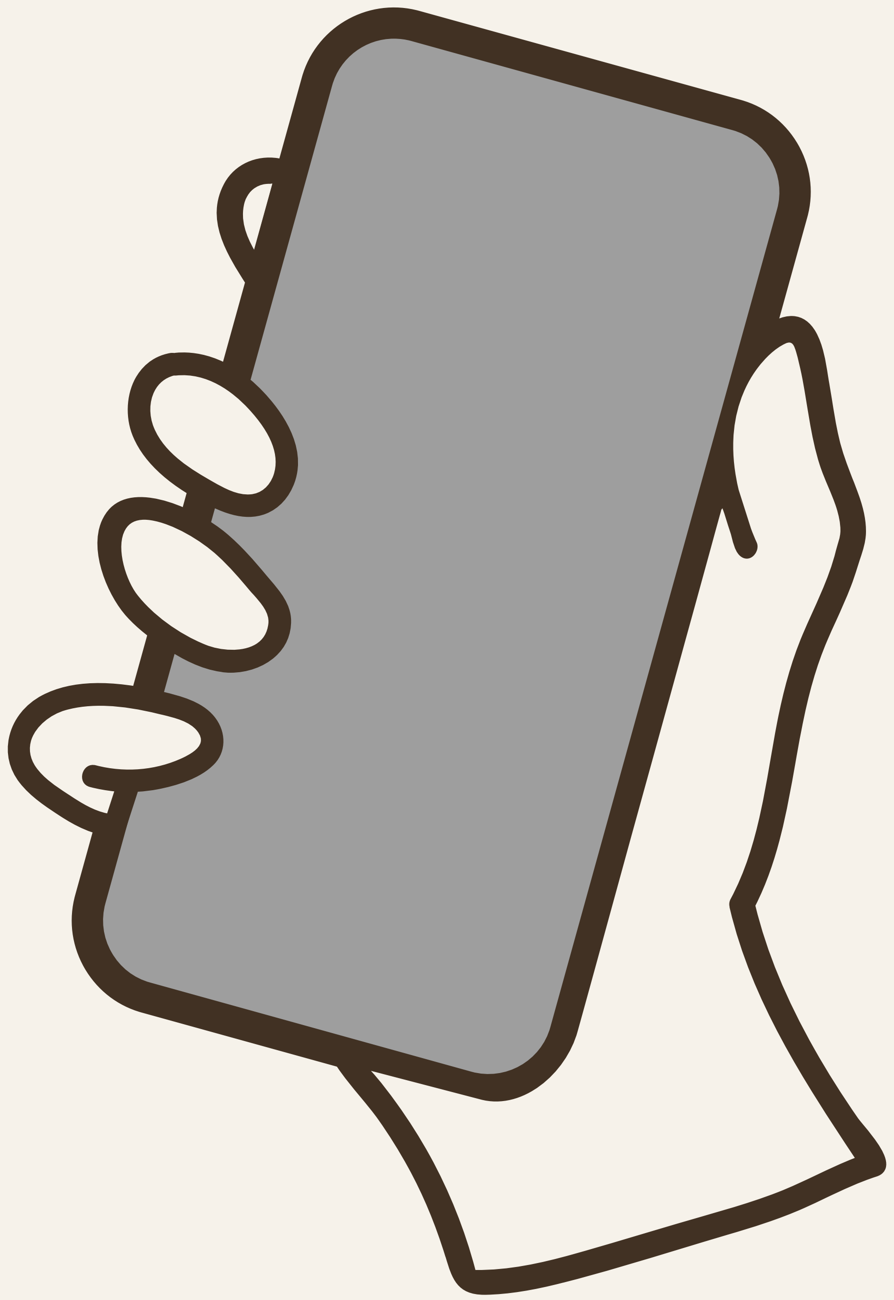
	precision	recall	f1-score	support
Bad	0.24	0.32	0.27	22
Normal	0.43	0.70	0.53	43
Good	0.00	0.00	0.00	35
accuracy			0.37	100
macro avg	0.22	0.34	0.27	100
weighted avg	0.24	0.37	0.29	100

PROSODIC

APPLICATION GUIDE FOR PARENTS

THROUGH MANY OF THE ABOVE PROCESSES,
WE HAVE HELPED CHILDREN LEARN MEANINGFULLY
AND EVALUATED THEM BASED ON DATA

IT'S TIME TO SHARE THE RESULTS OF LEARNING
WITH PARENTS TO MAKE YOUR CHILD'S LEARNING
MORE EFFICIENT



APPLICATION GUIDE FOR PARENTS

SCREEN THAT CONFIGURES THE
APP CONSISTS OF SIX TYPES:
LOGIN, HOME, STUDY LIST, AI REPORT,
CHALLENGE AND PROFILE

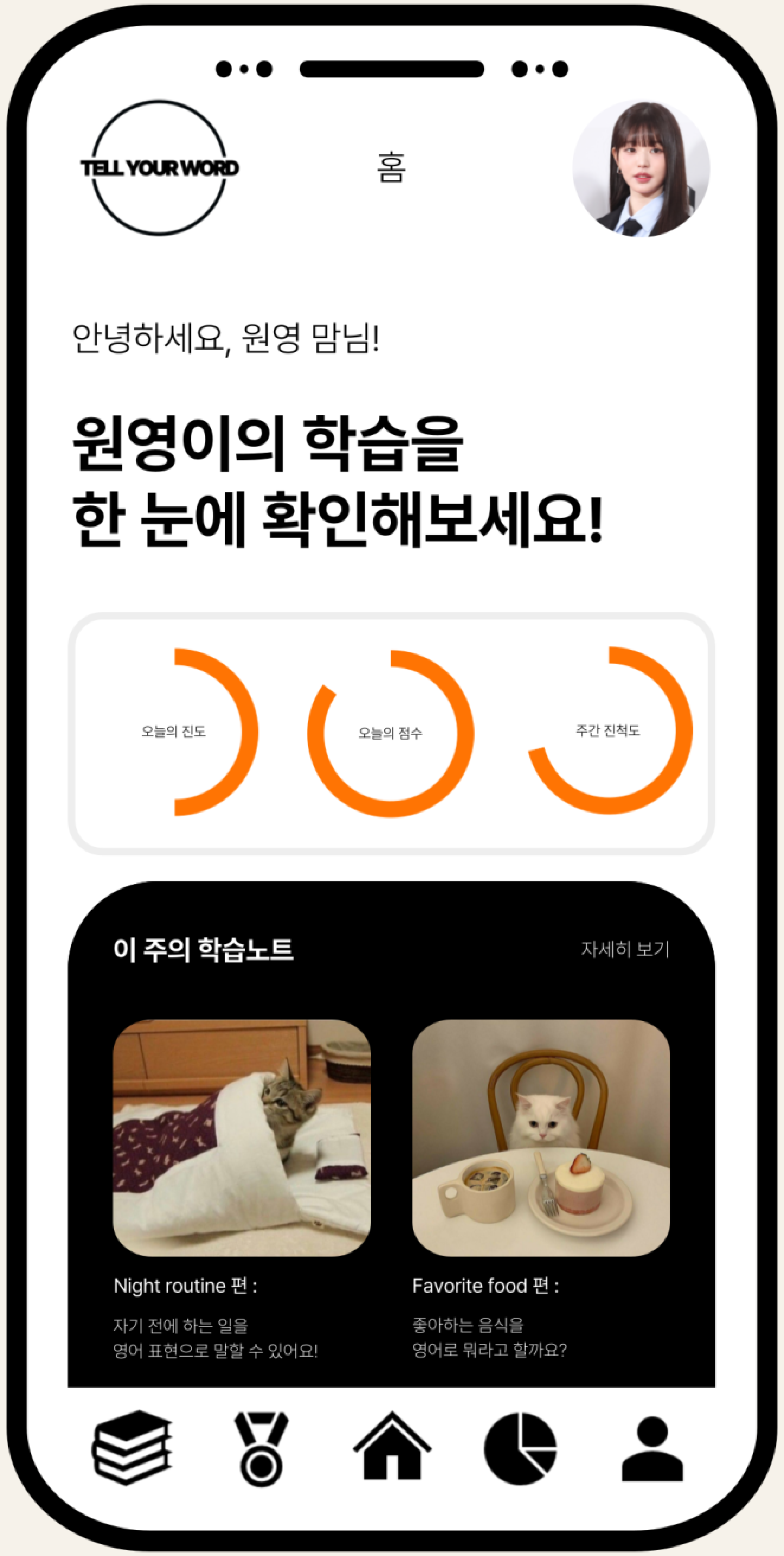
IN THE PRESENTATION,
WE WILL INTRODUCE SOME SCREENS
FOCUSING ON FUNCTIONS

HOME SCREEN

PROVIDE LEARNING DATA IN GRAPHS
SO PARENTS CAN CHECK THEIR CHILD'S
LEARNING STATUS AT A GLANCE

PROVIDE A SUMMARY OF WHAT THEIR CHILD
LEARNED THIS WEEK

IF PARENTS WANT TO KNOW MORE,
THEY CAN TOUCH AND GO TO
THE LEARNING LIST SCREEN



STUDY LIST SCREEN

PROVIDES SEARCH SERVICES BASED ON STUDY DATES OR KEYWORDS

HELP FACILITATE REVIEW BY PROVIDING WHAT THEIR CHILD HAS LEARNED IN A FILE

TO LEARN MORE ABOUT WHAT THEIR CHILD HAS LEARNED TODAY,

PARENTS CAN TOUCH THE LEARNING NOTES AND GO TO THE AI REPORT SCREEN

AI REPORT SCREEN

DIGITIZE THE CHILD'S PRONUNCIATION,
SCORE IT, AND VISUALIZE IT DIRECTLY
TO HELP PARENTS UNDERSTAND IT EASILY

PROVIDES DATA SUCH AS
VOICE SCORES BY INDICATOR,
LEVEL COMPARED TO THE SAME AGE GROUP,
AND INCREASE IN MONTHLY TOTAL SCORE

ADOPT A GRAPH TYPE THAT
REPRESENTS EACH CHARACTERISTIC WELL

COMMENT AT THE BOTTOM OF THE GRAPH
TO BRIEFLY EXPLAIN THE WEAKNESSES

CHALLANGE SCREEN

NO MATTER HOW WELL THE CONDITIONS FOR LEARNING ARE, IT WILL BE DIFFICULT TO PARTICIPATE STEADILY WITHOUT ENCOURAGING PARTICIPATION

MOTIVATION BY PROVIDING REWARDS
FOR LEARNING (EX) T MEMBERSHIP POINTS

REWARDED BY PERFORMING SIMPLE MISSIONS
OR CONTINUING TO LEARN

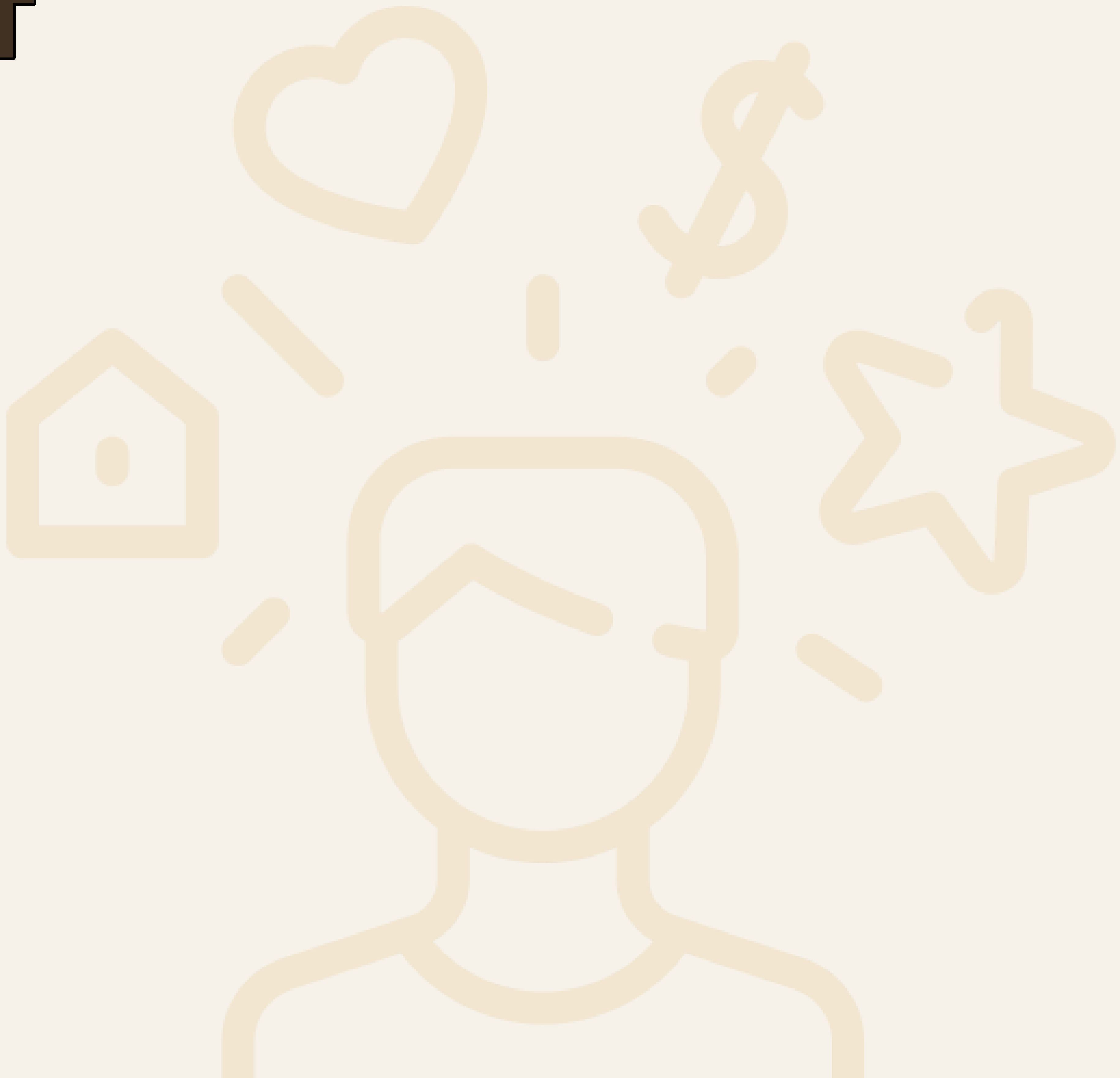
IT'S NOT JUST A SERVICE, BUT IT GIVES YOU
THE FEELING OF BEING A MEMBER OF YOUR CHILD'S GROWTH

EXPECTED ACHIEVEMENT

SUPPORTING CHILDREN TO IMPROVE
THEIR PRONUNCIATION SKILLS

OFFERING MORE ACTIVE
ENGLISH EDUCATION MATERIALS TO THE USERS

ANALYZING AND MAKING FUTURE WORK
ABOUT CHILDREN'S LANGUAGE DEVELOPMENT



LIMITATION

NEED MORE LABELLED DATA
FOR ACCURACY

NEED PROFESSIONAL ANNOTATORS
TO DETERMINE PRONUNCIATION SCORES

NEED TO CHECK IF NUGU SPEAKER
CAN SAVE SPEAKERS' .WAV DATA

THANK YOU FOR LISTENING

TELL YOUR WORDS

KIM JOONHEE, DEPT OF INFORMATION SYSTEM 18
MA YUDAM, DEPT OF KOREAN LANGUAGE & LITERATURE 19
PARK JUNHYEONG, DEPT OF BUSINESS ADMINISTRATION 18