Grapheme To Phoneme Converter

**(Amroz Siddiqui)**

**Joshua Philip [1021155] Assistant Professor**

**Atharv Kadam [1021156] Department of Computer Engineering Niranjan Kurhade [1021164] Fr.CRIT, Vashi**

# Abstract

The Grapheme to Phoneme Converter is a web-based application that converts written word strings into their phonetic representations using a grapheme dictionary, the International Phonetic Alphabet (IPA), and a

text-to-speech (TTS) API. This is tool for language learning which provides phonetic transcriptions and audio pronunciations.

# Introduction

Understanding the relationship between written and spoken language is essential in linguistics and language education. Graphemes represent written symbols, while phonemes are distinct sound units. The Grapheme to Phoneme Converter allows users to input word strings and receive phonetic transcriptions along with audio examples. By integrating a grapheme dictionary, IPA conversion, and a TTS API, this web application serves as a valuable resource for language learners, educators, and linguists, facilitating better pronunciation and comprehension.

# Analysis

The Grapheme to Phoneme Converter follows a straightforward process: it receives a word string, consults a grapheme dictionary to identify graphemes, and converts these to phonemes using the IPA. This standardized representation ensures clarity in phonetic transcription. The application then utilizes the Google Text-to-Speech (GTTS) API to generate audio files for each phoneme, allowing users to hear accurate pronunciations. The web interface displays IPA transcriptions alongside audio playback, creating an interactive learning experience that is accessible to users of different language proficiencies.

# Design

# Uml Sequence Diagram

# 

1. **Implementation**

Searching Input Word in CMU Dictionary

def s e a r c h w o r d i n f i l e ( f i l e p a t h , search word ) : with open ( f i l e p a t h , ’ r ’ ) as f i l e :

l i n e s = f i l e . r e a d l i n e s ( )

f o r l i n e in l i n e s :

# S p l i t the l i n e i n to the word and i t s n o ta t i o n

p a r ts = l i n e . s t r i p ( ) . s p l i t ( ’ *\* t ’ ) # Assuming the f i l e uses tab as d e l i m i t e r i f l e n ( p a r ts ) *>* 1 :

word = p a r ts [ 0 ]

n o ta t i o n = p a r ts [ 1 ]

i f word . lower ( ) == search word . lower ( ) : # Case i n s e n s i t i v e s earc h r etur n n o ta t i o n

r etur n None # Return None i f the word i s not found

Getting IPA notation

def get ipa and example ( arpabet symbol ) :

””” Re tr i e ve IPA n o ta t i o n and example word from ARPAbet symbol . ””” r etur n a r p a b e t t o i p a . get ( arpabet symbol . upper ( ) , None )

Main function

def main ( u s e r i n p u t ) :

search word = u s e r i n p u t

n o ta t i o n = s e a r c h w o r d i n f i l e ( f i l e p a t h , search word )

i f n o ta t i o n :

p h o n e t i c r e p r e s e n t a t i o n = n o ta t i o n . s p l i t ( ) n o ta t i o n i n d e x = 0

j s o n o u tp u t = *{}*

f o r symbol in p h o n e t i c r e p r e s e n t a t i o n :

r e s u l t = get ipa and example ( symbol ) i f r e s u l t :

# p r i n t ( f ”ARPAbet : *{* symbol *}* , IPA : *{* r e s u l t [ ’ IPA ’ ] *}* , Example : *{* r e s u l t [ ’ Example ’ ]

j s o n o u tp u t [ symbol ] = *{*

’ IPA ’ : r e s u l t [ ’ IPA ’ ] ,

’ Example ’ : r e s u l t [ ’ Example ’ ]

*}*

t t s = gTTS( te x t=r e s u l t [ ’ Example ’ ] , lang =’en ’ ) t t s . save ( f ” s t a t i c / audio /*{* symbol *}* . mp3”)

n o ta t i o n i n d e x+=1 r etur n j s o n o u tp u t

# Conclusion

The Grapheme to Phoneme Converter tries to bridge the gap between written and spoken language, offering accurate phonetic transcriptions and audio examples. This project can be a good resource for language learners and educators, enhancing pronunciation skills and comprehension. Future enhancements could include expanding the dictionary to support multiple languages and dialects, further increasing its utility. Overall, this project highlights the potential of technology to facilitate language learning and improve literacy.

# Acknowledgement

* 1. The CMU Pronouncing Dictionary <http://www.speech.cs.cmu.edu/cgi-bin/cmudict?in=hello>
  2. Wikipedia-ARPABET https://en.wikipedia.org/wiki/ARPABET
  3. Google Text-to-Speech https://gtts.readthedocs.io/en/