卒 業 論 文 概 要 書

Summary of Bachelor's Thesis

Date of submission: 07/24/2023 (MM/DD/YYYY)

学科名 Department	Major in Computer Science and Communications Engineering	氏名 Name	Touko Aoki	指　導 教　員 Advisor	Hiroshi Watanabe 印
研究指導名 Research guidance	Research on Audiovisual Information Processing	学籍番号 Student ID number	1W19CS01-3		
研究題目 Title	Evaluation of the Pronunciation Similarity Between English and Japanese Words Using the International Phonetic Alphabet Representation				

1. Introduction

Learning different languages can be a challenging attempt, presenting a multitude of difficulties to overcome. In particular, correctly acquiring unfamiliar pronunciations of foreign languages can be a great obstacle. In our research, we employed International Phonetic Alphabet (IPA) as the unified representation of multiple languages. Also, we proposed a method to quantify the distance between each unit of sound by numerical values. As a result, we successfully found words pronounced similarly in English and Japanese, even though they apply exceedingly variant characters in their own scripts.

2. Related Works

Soundex[1] is a phonetic algorithm for indexing and matching words based on their pronunciation, initially developed to reduce similar-sounding words to the same code, to efficiently search for names. Soundex has been updated with more advanced phonetic algorithms such as NYSIIS[2], and Metaphone[3] developed to enhance its performance. However, those methods have the critical limitation that they are only capable of evaluating Latin languages since they applied English alphabets for indexing words.

3. Proposed Method

3.1. International Phonetic Alphabet

In the experiment, we employed the International Phonetic Alphabet (IPA), provided by the International Phonetic Association, as the symbols to notate the pronunciations of words[4]. IPA features an extensive chart that represents the sounds of human speech. For consonants, the chart classifies them by two indices: place and manner. Similarly, vowels are categorized by openness and backness, and roundness. These indices indicate the articulation of the mouth during the pronunciation.

3.2. Euclidean Distance

To quantify the similarity between each IPA by the numerical values, we applied the Euclidean distance function. Euclidean distance[5] is the equation to define the straight-line distances between two points. For n-dimensional space, the Euclidean distance d between two points $A(x_1, x_2, \ldots, x_n)$ and $B(y_1, y_2, \ldots, y_n)$ is given by:

$$d = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2 + \cdots + (y_n - x_2)^2} \quad (1)$$

3.3. Normalization and Scaling

Setting the weights of individual phonetic features is the essential procedure to rationally measure the distances between each IPA. We normalized and scaled the distance function to achieve specific conditions in terms of the Euclidean distance between points. The requirements include the equalization of the contribution by inflating numerically small dimensions and deflating numerically large ones. We also fixed the maximum distance between any two points as 1.0, while an average distance between points is assumed as 0.5.

Then, we applied a non-linear scaling power, denoted as r, which allows for balancing the dimensions to raise the weights. The power r can be adjusted to approach the target value of 1.0 for the maximum distance. Then, weights are scaled by a uniform scaling factor k. Therefore, we can obtain w_i, the i-th weight, by equation 4 below.

$$w_i = k\left(\frac{1}{\sum_{x,y \in P}|x_i - y_i|}\right)^r \quad (4)$$

4. Experiments

4.1. Procedures

First, we classified the features of consonants and vowels based on the indices provided by the IPA chart. Then, we calculated the distances between each IPA by the Euclidean distance function. The obtained values were compiled as two tables for consonants and vowels. TABLE 1 represents the distance between each consonant while TABLE 2 is for vowels.

TABLE 1 Consonants distance table

	p	b	t	d	k	g	m	n	ŋ	ʁ	r	ɸ	f	v	θ	ð	s	z	ʃ	ʒ	ç	h	ɹ	j	ɰ	w	ɐ	ɚ	ɫ
p	0	0.01	0.02	0.03	0.08	0.08	0.12	0.14	0.17	0.18	0.21	0.23	0.25	0.26	0.27	0.28	0.29	0.3	0.31	0.32	0.34	0.42	0.46	0.48	0.49	0.64	0.75	0.76	1
b	0.01	0	0.02	0.02	0.07	0.08	0.11	0.13	0.16	0.17	0.21	0.22	0.24	0.25	0.26	0.27	0.28	0.29	0.3	0.31	0.34	0.41	0.46	0.48	0.49	0.64	0.75	0.76	1
t	0.02	0.02	0	0.01	0.06	0.07	0.11	0.12	0.15	0.16	0.2	0.22	0.24	0.25	0.26	0.26	0.27	0.28	0.29	0.3	0.33	0.41	0.45	0.47	0.48	0.63	0.75	0.75	1
d	0.03	0.02	0.01	0	0.05	0.06	0.1	0.12	0.14	0.15	0.19	0.21	0.23	0.24	0.25	0.26	0.27	0.28	0.29	0.3	0.32	0.4	0.45	0.47	0.48	0.63	0.74	0.75	1
k	0.08	0.07	0.06	0.05	0	0.01	0.05	0.07	0.1	0.11	0.15	0.17	0.19	0.2	0.21	0.22	0.23	0.24	0.25	0.26	0.29	0.37	0.42	0.44	0.45	0.61	0.73	0.74	1
g	0.08	0.08	0.07	0.06	0.01	0	0.05	0.06	0.09	0.1	0.14	0.16	0.18	0.19	0.2	0.21	0.22	0.23	0.24	0.25	0.28	0.36	0.41	0.43	0.44	0.61	0.73	0.74	1
m	0.12	0.11	0.11	0.1	0.05	0.05	0	0.03	0.05	0.06	0.11	0.13	0.15	0.16	0.17	0.18	0.19	0.2	0.21	0.22	0.25	0.34	0.39	0.41	0.42	0.59	0.72	0.73	1
n	0.14	0.13	0.12	0.12	0.07	0.06	0.03	0	0.04	0.05	0.09	0.11	0.13	0.14	0.15	0.16	0.17	0.18	0.19	0.2	0.24	0.32	0.38	0.4	0.41	0.58	0.71	0.72	1
ŋ	0.17	0.16	0.15	0.14	0.1	0.09	0.05	0.04	0	0.02	0.06	0.08	0.1	0.11	0.12	0.14	0.15	0.16	0.17	0.18	0.21	0.3	0.36	0.38	0.39	0.57	0.7	0.71	1
ʁ	0.18	0.17	0.16	0.15	0.11	0.1	0.06	0.05	0.02	0	0.05	0.07	0.09	0.1	0.11	0.13	0.14	0.15	0.16	0.17	0.2	0.29	0.35	0.37	0.38	0.56	0.7	0.71	1
r	0.21	0.21	0.2	0.19	0.15	0.14	0.11	0.09	0.06	0.05	0	0.03	0.05	0.06	0.07	0.08	0.1	0.11	0.12	0.13	0.17	0.26	0.32	0.34	0.35	0.54	0.68	0.69	1
ɸ	0.23	0.22	0.22	0.21	0.17	0.16	0.13	0.11	0.08	0.07	0.03	0	0.03	0.04	0.05	0.06	0.07	0.09	0.1	0.11	0.15	0.24	0.3	0.33	0.34	0.53	0.67	0.69	1
f	0.25	0.24	0.24	0.23	0.19	0.18	0.15	0.13	0.1	0.09	0.05	0.03	0	0.01	0.03	0.04	0.05	0.06	0.08	0.09	0.12	0.22	0.28	0.31	0.32	0.52	0.67	0.68	1
v	0.26	0.25	0.25	0.24	0.2	0.19	0.16	0.14	0.11	0.1	0.06	0.04	0.01	0	0.02	0.03	0.04	0.05	0.06	0.08	0.11	0.21	0.28	0.3	0.31	0.51	0.66	0.68	1
θ	0.27	0.26	0.26	0.25	0.21	0.2	0.17	0.15	0.12	0.11	0.07	0.05	0.03	0.02	0	0.01	0.03	0.04	0.05	0.06	0.1	0.2	0.27	0.29	0.3	0.51	0.66	0.67	1
ð	0.28	0.27	0.26	0.26	0.22	0.21	0.18	0.16	0.14	0.13	0.08	0.06	0.04	0.03	0.01	0	0.01	0.03	0.04	0.05	0.09	0.19	0.26	0.28	0.3	0.5	0.65	0.67	1
s	0.29	0.28	0.27	0.27	0.23	0.22	0.19	0.17	0.15	0.14	0.1	0.07	0.05	0.04	0.03	0.01	0	0.01	0.03	0.04	0.08	0.18	0.25	0.27	0.29	0.49	0.65	0.66	1
z	0.3	0.29	0.28	0.28	0.24	0.23	0.2	0.18	0.16	0.15	0.11	0.09	0.06	0.05	0.04	0.03	0.01	0	0.01	0.03	0.07	0.17	0.24	0.26	0.28	0.49	0.64	0.66	1
ʃ	0.31	0.3	0.29	0.29	0.25	0.24	0.21	0.19	0.17	0.16	0.12	0.1	0.08	0.06	0.05	0.04	0.03	0.01	0	0.01	0.06	0.16	0.23	0.25	0.27	0.48	0.64	0.65	1
ʒ	0.32	0.31	0.3	0.3	0.26	0.25	0.22	0.2	0.18	0.17	0.13	0.11	0.09	0.08	0.06	0.05	0.04	0.03	0.01	0	0.04	0.15	0.22	0.24	0.26	0.47	0.64	0.65	1
ç	0.34	0.34	0.33	0.32	0.29	0.28	0.25	0.24	0.21	0.2	0.17	0.15	0.12	0.11	0.1	0.09	0.08	0.07	0.06	0.04	0	0.11	0.19	0.21	0.23	0.45	0.62	0.63	1
h	0.42	0.41	0.41	0.4	0.37	0.36	0.34	0.32	0.3	0.29	0.26	0.24	0.22	0.21	0.2	0.19	0.18	0.17	0.16	0.15	0.11	0	0.09	0.12	0.13	0.38	0.57	0.59	1
ɹ	0.46	0.46	0.45	0.45	0.42	0.41	0.39	0.38	0.36	0.35	0.32	0.3	0.28	0.28	0.27	0.26	0.25	0.24	0.23	0.22	0.19	0.09	0	0.04	0.06	0.33	0.53	0.55	1
j	0.48	0.48	0.47	0.47	0.44	0.43	0.41	0.4	0.38	0.37	0.34	0.33	0.31	0.3	0.29	0.28	0.27	0.26	0.25	0.24	0.21	0.12	0.04	0	0.03	0.3	0.52	0.54	1
ɰ	0.49	0.49	0.48	0.48	0.45	0.44	0.42	0.41	0.39	0.38	0.35	0.34	0.32	0.31	0.3	0.3	0.29	0.28	0.27	0.26	0.23	0.13	0.06	0.03	0	0.29	0.51	0.53	1
w	0.64	0.64	0.63	0.63	0.61	0.61	0.59	0.58	0.57	0.56	0.54	0.53	0.52	0.51	0.51	0.5	0.49	0.49	0.48	0.47	0.45	0.38	0.33	0.3	0.29	0	0.17	0.19	0.57
ɐ	0.75	0.75	0.75	0.74	0.73	0.73	0.72	0.71	0.7	0.7	0.68	0.67	0.67	0.66	0.66	0.65	0.65	0.64	0.64	0.64	0.62	0.57	0.53	0.52	0.51	0.17	0	0.01	0.33
ɚ	0.76	0.76	0.75	0.75	0.74	0.74	0.73	0.72	0.71	0.71	0.69	0.69	0.68	0.68	0.67	0.67	0.66	0.66	0.65	0.65	0.63	0.59	0.55	0.54	0.53	0.19	0.01	0	0.32
ɫ	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0.57	0.33	0.32	0

TABLE 2 Vowels distance table

	i	e	ɛ	a	ɑ	ɔ	o	u	ɯ	ɪ	ʊ	ə	æ
i	0	0.1	0.12	0.14	0.2	0.24	0.28	0.3	0.63	0.75	0.84	0.88	1
e	0.1	0	0.11	0.12	0.18	0.22	0.24	0.29	0.62	0.74	0.83	0.87	1
ɛ	0.12	0.11	0	0.1	0.16	0.17	0.23	0.26	0.6	0.73	0.83	0.87	1
a	0.14	0.12	0.1	0	0.11	0.17	0.2	0.23	0.58	0.72	0.82	0.86	1
ɑ	0.2	0.18	0.16	0.11	0	0.11	0.14	0.16	0.55	0.71	0.8	0.85	1
ɔ	0.24	0.22	0.17	0.17	0.11	0	0.14	0.15	0.53	0.69	0.79	0.84	1
o	0.28	0.24	0.23	0.2	0.14	0.14	0	0.13	0.51	0.68	0.78	0.83	1
u	0.3	0.29	0.26	0.23	0.16	0.15	0.13	0	0.47	0.66	0.76	0.82	1
ɯ	0.63	0.62	0.6	0.58	0.55	0.53	0.51	0.47	0	0.26	0.36	0.42	0.62
ɪ	0.75	0.74	0.73	0.72	0.71	0.69	0.68	0.66	0.26	0	0.16	0.22	0.35
ʊ	0.84	0.83	0.83	0.82	0.8	0.79	0.78	0.76	0.36	0.16	0	0.12	0.24
ə	0.88	0.87	0.87	0.86	0.85	0.84	0.83	0.82	0.42	0.22	0.12	0	0.19
æ	1	1	1	1	1	1	1	1	0.62	0.35	0.24	0.19	0

4.2. Results

The total distance between words is calculated by summing individual distances between each IPA. Fig. 1 shows the 10 pairs of English-Japanese words with the highest similarity of consonants and Fig. 2 displays the 10 pairs having similar vowels.

```
English bbc, Japanese: ＥＣＰ, Distance: 0.0
English bbc, Japanese: いき値, Distance: 0.0
English bbc, Japanese: いりちー, Distance: 0.0
English bbc, Japanese: きしり, Distance: 0.0
English bbc, Japanese: きりり, Distance: 0.0
English bbc, Japanese: ぎっしり, Distance: 0.0
English bbc, Japanese: ひひいん, Distance: 0.0
English bbc, Japanese: びしり, Distance: 0.0
English bbc, Japanese: びっしり, Distance: 0.0
English bbc, Japanese: びびり, Distance: 0.0
```

Fig. 1. 10 pairs of words with consonant similarities.

```
English abacus, Japanese: ＶＡＸ, Distance: 0.0
English abacus, Japanese: ばこそ, Distance: 0.0
English abacus, Japanese: バッカス, Distance: 0.0
English abacus, Japanese: バックス, Distance: 0.0
English abacus, Japanese: バックソー, Distance: 0.0
English abacus, Japanese: ボクサー, Distance: 0.0
English abacus, Japanese: ボックス, Distance: 0.0
English abacus, Japanese: 化かす, Distance: 0.0
English abacus, Japanese: 魅す, Distance: 0.0
English abacus, Japanese: 暈かす, Distance: 0.0
```

Fig. 2. 10 pairs of words with vowel similarities.

To verify the performance accuracy of the program and the processed distance tables, we then calculated the distances between the overall IPA, considering the order of consonants and vowels as well. Fig. 3 is for the top 10 pairs of words with the closest average distance between the overall IPA.

```
English minniti, Japanese: 耳ピ, Distance: 0.01
English ddt, Japanese: ビビビッ, Distance: 0.01
English ddt, Japanese: ピーピーピー, Distance: 0.01
English bbc, Japanese: びびり, Distance: 0.02
English bbc, Japanese: ビビリ, Distance: 0.02
English siddiqi, Japanese: 響, Distance: 0.02
English gdp, Japanese: ビビビッ, Distance: 0.02
English gdp, Japanese: ピーピーピー, Distance: 0.02
English gdp, Japanese: 忌引, Distance: 0.02
English gdp, Japanese: 忌引き, Distance: 0.02
```

Fig.3. 10 pairs of words with comprehensive distances.

5. Conclusion

In this experiment, we successfully quantified the distances between each IPA symbol by the Euclidean distance function. We also found words containing similar phonetic features from English and Japanese. The greatest contribution of this research was that we proposed a method to compare pronunciations by objective measurements, even across languages with variant types of characters. However, we still had some limitations in the experiment, such as the restriction of the dataset and the omission of a human evaluation survey to support our hypothesis. We strongly believe that future research can consider the variety of evaluation methods for giving a solution to potential challenges in acquiring unfamiliar pronunciation in foreign languages.

References

[1] David Odden, *Introducing phonology*. Cambridge university press, 2005. (Online ISBN: 9780511808869, https://doi.org/10.1017/CBO9780511808869)

[2] Jeffrey M. Silbert, "The World's First Computerized Criminal-Justice Information-Sharing System-The New York State Identification and Intelligence System (NYSIIS)." *Criminology*, Vol.8, No.2, pp.107-213, Aug. 1970. (https://doi.org/10.1111/j.1745-9125.1970.tb00734.x)

[3] Lawrence Philips, "Hanging on the metaphone." *Computer Language* Vol.7, No.12, pp.39-44, Dec. 1990.

[4] *Handbook of the International Phonetic Association: a Guide to the Use of the International Phonetic Alphabet*. Cambridge University Press, 1999. (ISBN: 0521652367)

[5] Verma, Dhruv, and S. N. Muralikrishna. "Semantic similarity between short paragraphs using Deep Learning." *2020 IEEE International Conference on Electronics, Computing and Communication Technologies (CONECCT)*. IEEE, pp.1-5, Jul. 2020.

Evaluation of the Pronunciation Similarity Between English and Japanese Words Using the International Phonetic Alphabet Representation

A Thesis Submitted to the Major in Computer Science and

Communications Engineering, the School of Fundamental Science and Engineering of

Waseda University

in Partial Fulfillment of the Requirements for the Degree of Bachelor of Engineering

July 24th, 2023

Touko Aoki

(1W19CS01-3)

Advisor: Prof. Hiroshi Watanabe

Research guideline: Research on Audiovisual Information Processing

Acknowledgments

I am profoundly thankful to Professor Hiroshi Watanabe, for providing me with a conducive research environment for my research. He always proposed significant insights to my experimental proposal and introduced me to numerous relevant studies to enhance the quality of my experiments.

Secondly, I would like to express my gratitude to Yenan Zhang, Guilly Kolkman, and Minghao Duan, the members of the 'audio seminar' at our laboratory. They provoked active discussions during the seminar and gave me thoughtful advice on evaluating my experimental methods. Their constructive feedback significantly contributed to the robustness of my findings. Additionally, their research in audio processing was so intriguing that they brought a bunch of great inspirations to my work as well.

Furthermore, I extend my thanks to all the members of my laboratory. They assisted me in setting up the experimental environment and also provided me with thoughtful questions during our seminars. The stimulating discussions within the laboratory have played an instrumental role in exploring my research ideas.

Lastly, I owe an immense debt of gratitude to my parents. Their financial support enabled me to pursue my research at Waseda University. Also, their selfless dedication to my education and personal growth has been the foundation of my academic journey.

Abstract

One of the major challenges in acquiring a new language is pronunciation. Particularly, reproducing sounds that do not exist in one's native language can be exceedingly difficult. In our research, we propose an algorithm to find pairs of words with similar pronunciations from multiple languages. There are some established methods such as Soundex and Metaphone in related studies to compare language similarities. However, these methods use general alphabets for sound classification, limiting their functionality to Latin-based languages.

In our approach, we first convert words written in their actual language script into the International Phonetic Alphabet (IPA), enabling a standardized comparison of all languages. We then analyze the phonetic characteristics of each IPA using an IPA chart to calculate the distances between them. By calculating each Euclidean distance, we generated separate tables for consonant and vowel correspondence. The average of each distance between the IPA symbols used in each word allowed us to obtain the overall distance between the words.

List of Contents

List of Figures

List of Tables

1. Introduction

1.1. Background

Learning different languages can be a challenging attempt, presenting a multitude of difficulties to overcome. There is a variety of languages having individual grammatical structures, vocabulary, and pronunciation. Acquiring such rules and patterns can be overwhelming for learners. Particularly, pronunciation can be a great obstacle among those aspects. There are various sounds and intonations in languages that are vastly different or do not exist in one's native language. Those sounds might be quite difficult for learners to master, as they need to work on the coordination of muscles and articulatory organs in new and unfamiliar ways. There is also a significant issue in the written form of language. Some languages have unique logographic characters or irregular pronunciation rules for each letter, which make it more challenging for learners to acquire correct pronunciation from its script.

Contemplating this issue, former researchers have developed the algorithms such as Soundex[1] and Metaphone[2] to analyze phonetic features by written language. However, those methods still have critical limitations in that they are only capable of evaluating Latin languages because they applied English alphabets for indexing words. As the solution for this, We employed the International Phonetic Alphabet (IPA) as the key index of pronunciation. Using the IPA, we aimed to visually analyze the distinct phonetic characteristics of languages, even if they have entirely different writing systems.

1.2. Objectives

The primary objective of our research is to conduct a comprehensive evaluation of languages having different types of characters, focusing on their phonetic aspects. We have chosen English and Japanese as the languages for comparison in our research. To achieve our goal, the International Phonetic Alphabet (IPA) was employed as a tool for representing and analyzing the sounds of speech. Using the IPA, we aim to visually and systematically explore phonetic differences and similarities between languages. Through our research, we seek to shed light on the specific phonetic features of each language, particularly examining the factors of vowel and consonant articulation.

1.3. Outline of thesis

The thesis paper consists of the following contents:

1. The first chapter serves as an introduction, defining our research theme (comparing the phonetic characteristics of different languages), and explaining the research background and motivation. Our study primarily focuses on using the International Phonetic Alphabet (IPA) to identify phonetic differences in vocabularies across multiple languages and proposing methods to assist learners in acquiring foreign language pronunciation.
2. The second chapter introduces two related prior studies to our research. Soundex is one of the most classical algorithms established with the purpose of systematically classifying languages phonetically and comparing their features. Metaphone, based on Soundex's theory, improved it to consider more linguistic features. It is used for services such as detecting misspellings and organizing indexes from lists containing names.

3. In the following chapter, we introduced the proposed method in the experiment. The relationship between IPA and IPA numbers is explained in detail in the first paragraph, and the characteristics of consonants and vowels are defined according to the 'IPA chart', provided by the International Phonetic Association. We also define the Euclidean distance function for measuring distances between IPAs and the formulas for normalizing the calculation results.
4. The experimental procedures and results are presented. First, we classified consonants and vowels by converting IPA symbols to IPA numbers and created separate feature tables for each consonant and vowel from the phonetic characteristics provided in the IPA chart. One of the most crucial processes was that we generated a table representing the distances between each IPAs using the Euclidean distance function. As experimental results, we identified pairs of words from English and Japanese word lists that contain similar vowel features and similar consonant features. The program outputs their similarity to the word pairs.
5. The section provides a discussion of the experiment. We compared the obtained data and evaluate the experimental accuracy and mentioned the remaining challenges revealed in the experiment and functional constraints in the experimental design. Additionally, we propose points of improvement and suggest future research themes.

2. Related Works

Unfortunately, there are still not enough research benchmarks or established methods to compare the vocabularies of languages with significantly different phonetic features (e.g., English and Japanese). Therefore, we got inspiration from two prior studies Soundex and Metaphone, which are commonly used for comparing English and European languages both using Latin characters.

2.1. Soundex

Soundex is a phonetic algorithm first patented in 1918 by Robert C. Russell and Margaret King Odell[1], for indexing and matching words based on their pronunciation. It was initially developed to reduce similar-sounding words to the same code, which enables efficient searching and matching for names that may have different spellings but sound similar when pronounced. The Soundex algorithm works as follows:

1. Convert the word to uppercase.
2. Keep the word's first letter as it is and drop all occurrences of the letters 'a', 'e', 'i', 'o', 'u', 'w', 'h', and 'y'.
3. Assign numeric codes to the remaining letters. The classification of Soundex phonetic codes is given in TABLE 1.1[2].

TABLE 1.1 Soundex phonetic codes

Letters	Assigned Code
a, e, i, o, u, w, h, y	0
b, f, p, v	1
c, g, j, k, q, s, x, z	2
d, t	3
l	4
m, n	5
r	6

4. Combine the first letter with the numeric codes obtained in step 3, up to a total of four characters. If the word doesn't have enough characters, pad with zeros.
5. The resulting Soundex code is the combination of the first letter and the three numeric codes (e.g., the word "Smith" would become "S530").

There have been more advanced phonetic algorithms such as NYSIIS[3], and Metaphone[4] developed to address some of the limitations of classic Soundex.

2.2. Metaphone

Metaphone was the updated algorithm performed by Lawrence Philips in 1990 as an improvement over Soundex mentioned previously. The main objective of Metaphone is to produce more accurate phonetic representations of words, especially when dealing with a wide range of names and words from various languages and origins. Unlike Soundex, which produces a fixed-length code of up to four characters, Metaphone generates variable-length codes, usually shorter than Soundex codes. The algorithm considers various English pronunciation rules and patterns to create a more reliable phonetic representation of a word. The Metaphone algorithm works as follows:

1. Convert the word to uppercase.
2. Remove non-alphabetic characters and consecutive duplicate letters.
3. Process special cases and transformations based on English pronunciation.
4. Encode the remaining letters into a phonetic code based on their pronunciation patterns.

Metaphone takes into account a variety of rules and transformations, such as handling silent letters, certain letter combinations (e.g., "ph" usually becomes "F"), and the sound of certain letter pairs (e.g., "th" is replaced by "T" or "0" depending on its pronunciation). The resulting code is a phonetic representation of the word that captures its pronunciation more accurately. TABLE 1.2 below shows the encodings of words and the Metaphone library[5].

TABLE 1.2 Example of Metaphone encoding

Word Variants	**CMU Expanded Pronunciation**	**Metaphone Encoding**
PROZAC*	***P R OW Z AE K***	***PROZACPRSK***
POZAC	P AA Z AH K	POZACPSK
PRZAC	P R Z AE K	**PRZACPRSK**
PROAC	P R OW AE K	PROACPRK
PROZAK	**P R OW Z AE K**	**PROZAKPRSK**
PROXAC	**P R OW Z AE K**	PROXACPRKS

*correct spelling of the word

Metaphone has been widely used in applications such as database indexing, record matching, and information retrieval where accurate matching of words based on their pronunciation is essential. It has also inspired variations like Double Metaphone[6], which produces both a primary and an alternate code, and Metaphone3[7], which is an enhanced version designed to handle a broader range of languages and names.

3. Proposed method

3.1. IPA symbols to IPA number conversion

Our proposed methodology involves the conversion of IPA (International Phonetic Alphabet) symbols into IPA numbers, originally provided by the Kiel Convention in 1989 and revised by the International Phonetic Association in 2020[8]. IPA numbers are 3-digit codes assigned to individual IPA symbols. These numbers were originally introduced to explicitly identify IPA symbols during the era of competing computer encodings, thereby preventing confusion between similar characters when dealing with manuscript printing.

The IPA numbers are structured based on the magnitude of the digits, allowing for distinct categorization of symbols to classify their semantic and graphic characteristic. Consonants are represented in the 100 series, while retired and non-IPA consonants occupy the 200s. Vowels are designated in the 300s, diacritics in the 400s, suprasegmentals in the 500s, extIPA symbols in the 600s, capital letters in the 700s, and transcription delimiters in the 900s. Some symbols may have multiple codes assigned to them.

By converting IPA symbols to IPA numbers, each numerical value serves as an index representing the phonetic characteristics of the corresponding IPA symbol. This enables us to effectively categorize consonants and vowels within our program and allows for efficient handling of linguistic characters.

3.2. IPA features

The International Phonetic Alphabet features an extensive chart that represents the sounds of human speech. For consonants, the IPA chart classifies them by place and manner of articulation. Places of articulation include bilabial, labiodental, dental, alveolar, palatal, velar, and glottal. Manners of articulation encompass plosives, nasals, trills, tap or flaps, fricatives, lateral fricatives, approximants, and lateral approximants. These features are summarized in Fig. 3.1 below ("Full IPA Chart" is provided by International Phonetic Association[8]).

	Bilabial	Labiodental	Dental	Alveolar	Postalveolar	Retroflex	Palatal	Velar	Uvular	Pharyngeal	Glottal
Plosive	p b			t d		ʈ ɖ	c ɟ	k g	q ɢ		ʔ
Nasal	m	ɱ		n		ɳ	ɲ	ŋ	ɴ		
Trill	ʙ			r					ʀ		
Tap or Flap		ⱱ		ɾ		ɽ					
Fricative	ɸ β	f v	θ ð	s z	ʃ ʒ	ʂ ʐ	ç ʝ	x ɣ	χ ʁ	ħ ʕ	h ɦ
Lateral fricative				ɬ ɮ							
Approximant		ʋ		ɹ		ɻ	j	ɰ			
Lateral approximant				l		ɭ	ʎ	ʟ			

Symbols to the right in a cell are voiced, to the left are voiceless. Shaded areas denote articulations judged impossible.

Fig. 3.1. IPA consonant chart.

As for vowels, the IPA chart categorizes them based on tongue height and backness, ranging from high to low and front to back respectively. There are also distinctions between rounded and unrounded vowels. Similarly to the IPA consonants chart, these features of consonants are compiled in Fig. 3.2[8].

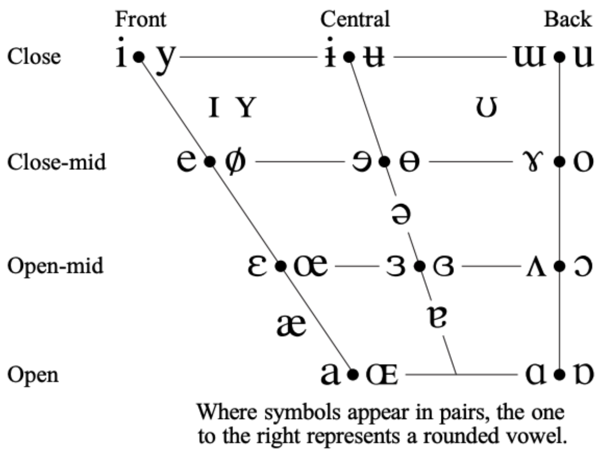

Fig. 3.2. IPA vowel chart.

3.3. Euclidean Distance

We applied the method of Euclidean distance to calculate the distances between each IPA according to their phonetic features. Euclidean distance is a fundamental concept in mathematics and is commonly used in various fields, including geometry, statistics, and machine learning. It measures the straight-line or "as-the-crow-flies" distance between two points in Euclidean space[9]. The Euclidean distance formula is derived from the Pythagorean theorem and provides a straightforward method to calculate the distance between two points in a Cartesian coordinate system.

In a two-dimensional Euclidean space, consider two points: A(x_1, y_1) and B(x_2, y_2). The Euclidean distance between these two points is given by the formula:

$$d = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2} \quad (2.1)$$

This formula represents the square root of the sum of the squares of the differences between the coordinates of the two points. It measures the length of the straight line segment connecting the two points. The Euclidean distance formula can be generalized to higher dimensions as well, where each additional dimension adds another term to the formula. For n-dimensional space, the Euclidean distance d between two points $A(x_1, x_2, ..., x_n)$ and $B(y_1, y_2, ..., y_n)$ is given by:

$$d = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2 + ... + (y_n - x_n)^2} \quad (2.2)$$

The Euclidean distance is always non-negative and satisfies the properties of a distance metric, such as symmetry ($d(A, B) = d(B, A)$) and the triangle inequality ($d(A, B) + d(B, C) \geq d(A, C)$). It provides a useful measure of similarity or dissimilarity between points in Euclidean space and serves as a foundation for various applications, including data analysis, pattern recognition, image processing, and more. Its simplicity and intuitive interpretation make it a popular choice in many mathematical and computational contexts.

3.4. Normalization and Scaling

In our experiment, we applied the normalization and scaling procedures to obtain the weights of individual phonetic features. Distance normalization and scaling aim to adjust weights to achieve specific conditions in terms of the Euclidean distance between points. The requirements include equalizing the contribution of each dimension, inflating numerically small dimensions and deflating numerically large ones, ensuring the maximum distance between any two points is 1.0, and obtaining an average distance between points of 0.5. These conditions are formally denoted by the following statements:

1. $\Sigma_{x,y\in P} w_1 \left|x_1 - y_1\right| \approx \Sigma_{x,y\in P} w_2 \left|x_2 - y_2\right| \approx \ \dots \ \approx \Sigma_{x,y\in P} w_n \left|x_n - y_n\right|$ (2.3)
2. $max_{x,y\in P} d(x,y) = 1.0$ (2.4)
3. $\Sigma_{x,y\in P} \frac{d(x,y)}{m^2} = 0.5$ (2.5)

To achieve these conditions, a non-linear monotonic function is applied to the weights. The weights are raised to a power, denoted as r, which allows for balancing the dimensions. The equation for calculating the weights involves dividing 1 by the sum of absolute differences between corresponding coordinates of points.

$$\Sigma_{x,y\in P} w_1 \left|x_1 - y_1\right| = \ \dots \ = \Sigma_{x,y\in P} w_n \left|x_n - y_n\right| = 1$$

$$\therefore w_1 \Sigma_{x,y\in P} \left|x_1 - y_1\right| = \ \dots \ = w_n \Sigma_{x,y\in P} \left|x_n - y_n\right| = 1$$

$$\therefore w_i = \frac{1}{\Sigma_{x,y \in P}\left|x_i - y_i\right|}$$

Apply r (non-linear scaling power) and then we get:

$$w_i = \left(\frac{1}{\Sigma_{x,y \in P}\left|x_i - y_i\right|}\right)^r.$$

Applying k (uniform scaling factor) gives:

$$w_i = k\left(\frac{1}{\Sigma_{x,y \in P}\left|x_i - y_i\right|}\right)^r. \tag{2.6}$$

Here, we can represent the equation above as the following form as well:

$$w_i = k\left(\frac{1}{\Delta_i}\right)^r, \tag{2.7}$$

where

$$\Delta_i = \Sigma_{x,y \in P}\left|x_i - y_i\right|.$$

Then, considering equation (2.5), the scaling factor k is determined to ensure that the average distance is 0.5. This is achieved by iterative computing k based on the sum of the adjusted distances and the desired average.

$$avg = \frac{\Sigma_{x,y \in P} d(x,y)}{m^2} = 0.5$$

$$\therefore \frac{\Sigma_{x,y \in P} k\sqrt{\left(\frac{x_1 - y_1}{\Delta_1^{\,r}}\right)^2 + \ldots + \left(\frac{x_n - y_n}{\Delta_n^{\,r}}\right)^2}}{m^2} = 0.5$$

$$\therefore k\Sigma_{x,y \in P}\sqrt{\left(\frac{x_1 - y_1}{\Delta_1^{\,r}}\right)^2 + \ldots + \left(\frac{x_n - y_n}{\Delta_n^{\,r}}\right)^2} = 0.5m^2$$

$$\therefore k = \frac{0.5m^2}{\sum_{x,y \in P} \sqrt{\left(\frac{x_1 - y_1}{\Delta_1^r}\right)^2 + \ldots + \left(\frac{x_n - y_n}{\Delta_n^r}\right)^2}} \tag{2.8}$$

The power r can be adjusted to approach the target value of 1.0 for the maximum distance. Repeated iterations using a binary search approach allow for the determination of r the desired level of accuracy. Once r is determined, the final weights can be calculated using the known scaling factor k. Substituting the adjusted weights back into the original distance equation gives the Euclidean distance between two points can be calculated, incorporating the scaled weights and the power r.

$$d(x, y) = \sqrt{\left[\frac{k(x_1 - y_1)}{\Delta_1^r}\right]^2 + \ldots + \left[\frac{k(x_n - y_n)}{\Delta_n^r}\right]^2}$$

$$\therefore d(x, y) = k\sqrt{\left(\frac{x_1 - y_1}{\Delta_1^r}\right)^2 + \ldots + \left(\frac{x_n - y_n}{\Delta_n^r}\right)^2} \tag{2.9}$$

4. Experiments

4.1. Procedures

To reduce the computational cost of the program, we extract only the IPA (International Phonetic Alphabet) symbols that are used in either English or Japanese word lists. Then, extracted IPA symbols are converted into IPA numbers and categorized into consonants and vowels based on their digits (3.1). We compiled the phonetic features of each IPA into the tables for consonants and vowels, based on the indices provided by the IPA chart. For consonants, we consider two indices:

1. Place: Describes the configuration of the lips during pronunciation
2. Manner: Represents the airflow and the obstruction,

whereas three indices are defined for vowels:

1. Openness: Indicates how much the mouth is open during pronunciation
2. Backness: Refers to the position of the tongue in the mouth during pronunciation (how far back or front it is)
3. Roundness: Indicates whether the lips are rounded during pronunciation.

TABLE 4.1 and TABLE 4.2 below summarize the phonetic features of consonants and vowels, respectively.

TABLE 4.1 Phonetic features of consonants

		p	b	t	d	k	g	m	n	ŋ	ɴ	ɾ	ɸ	f	v	θ	ð	s	z	ʃ	ʒ	ç	h	ɹ	j	ɰ	w	ɕ	ʑ	ɫ
Place	Bilabial	1	1	0	0	0	0	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	Labiodental	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	Alveolar	0	0	1	1	0	0	0	1	0	0	1	0	0	0	1	1	1	1	1	1	0	0	1	0	0	0	0	0	0
	Palatal	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	1	0	0	0	0	0
	Velar	0	0	0	0	1	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0
	Uvular	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	Glottal	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
Manner	Plosive	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	Nasal	0	0	0	0	0	0	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	Tap or Flap	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	Fricative	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0
	Approximant	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	0	0	0	0

TABLE 4.2 Phonetic features of vowels

		i	e	ɛ	a	ɑ	ɔ	o	u	ɯ	ɪ	ʊ	ə	æ
Openness	Close	1	0	0	0	0	0	0	1	1	0	0	0	0
	Near-close	0	0	0	0	0	0	0	0	0	1	1	0	0
	Mid-close	0	1	0	0	0	0	1	0	0	0	0	0	0
	Mid	0	0	0	0	0	0	0	0	0	0	0	1	0
	Mid-open	0	0	1	0	0	1	0	0	0	0	0	0	0
	Near-open	0	0	0	0	0	0	0	0	0	0	0	0	1
	Open	0	0	0	1	1	0	0	0	0	0	0	0	0
Backness	Front	1	1	1	1	0	0	0	0	0	1	0	0	1
	Central	0	0	0	0	0	0	0	0	0	0	0	1	0
	Back	0	0	0	0	1	1	1	1	1	0	1	0	0
	Rounded	0	0	0	0	1	1	1	1	0	0	1	0	0

Then, we applied the normalization and scaling procedures (mentioned in section 3.4) to assign weights to individual features shown in each row of the two tables. Applying those calculated weights gives the distances between each pair of IPA symbols by the Euclidean distance function, defined in section 3.3. The distances between IPA pairs are then normalized to decimal values between 0 and 1, where smaller values indicate that the corresponding IPA pair is phonetically closer. The following two tables, TABLE 4.3 and 4.4, represent the distance tables between IPA symbols for consonants and vowels, respectively.

TABLE 4.3 Consonants distance table

	p	b	t	d	k	g	m	n	ŋ	ɴ	r	ɸ	f	v	θ	ð	s	z	ʃ	ʒ	ç	h	ɹ	j	ɰ	w	ɕ	ʑ	ɬ
p	0	0.01	0.02	0.03	0.08	0.08	0.12	0.14	0.17	0.18	0.21	0.23	0.25	0.26	0.27	0.28	0.29	0.3	0.31	0.32	0.34	0.42	0.46	0.48	0.49	0.64	0.75	0.76	1
b	0.01	0	0.02	0.02	0.07	0.08	0.11	0.13	0.16	0.17	0.21	0.22	0.24	0.25	0.26	0.27	0.28	0.29	0.3	0.31	0.34	0.41	0.46	0.48	0.49	0.64	0.75	0.76	1
t	0.02	0.02	0	0.01	0.06	0.07	0.11	0.12	0.15	0.16	0.2	0.22	0.24	0.25	0.26	0.26	0.27	0.28	0.29	0.3	0.33	0.41	0.45	0.47	0.48	0.63	0.75	0.75	1
d	0.03	0.02	0.01	0	0.05	0.06	0.1	0.12	0.14	0.15	0.19	0.21	0.23	0.24	0.25	0.26	0.27	0.28	0.29	0.3	0.32	0.4	0.45	0.47	0.48	0.63	0.74	0.75	1
k	0.08	0.07	0.06	0.05	0	0.01	0.05	0.07	0.1	0.11	0.15	0.17	0.19	0.2	0.21	0.22	0.23	0.24	0.25	0.26	0.29	0.37	0.42	0.44	0.45	0.61	0.73	0.74	1
g	0.08	0.08	0.07	0.06	0.01	0	0.05	0.06	0.09	0.1	0.14	0.16	0.18	0.19	0.2	0.21	0.22	0.23	0.24	0.25	0.28	0.36	0.41	0.43	0.44	0.61	0.73	0.74	1
m	0.12	0.11	0.11	0.1	0.05	0.05	0	0.03	0.05	0.06	0.11	0.13	0.15	0.16	0.17	0.18	0.19	0.2	0.21	0.22	0.25	0.34	0.39	0.41	0.42	0.59	0.72	0.73	1
n	0.14	0.13	0.12	0.12	0.07	0.06	0.03	0	0.04	0.05	0.09	0.11	0.13	0.14	0.15	0.16	0.17	0.18	0.19	0.2	0.24	0.32	0.38	0.4	0.41	0.58	0.71	0.72	1
ŋ	0.17	0.16	0.15	0.14	0.1	0.09	0.05	0.04	0	0.02	0.06	0.08	0.1	0.11	0.12	0.14	0.15	0.16	0.17	0.18	0.21	0.3	0.36	0.38	0.39	0.57	0.7	0.71	1
ɴ	0.18	0.17	0.16	0.15	0.11	0.1	0.06	0.05	0.02	0	0.05	0.07	0.09	0.1	0.11	0.13	0.14	0.15	0.16	0.17	0.2	0.29	0.35	0.37	0.38	0.56	0.7	0.71	1
r	0.21	0.21	0.2	0.19	0.15	0.14	0.11	0.09	0.06	0.05	0	0.03	0.05	0.06	0.07	0.08	0.1	0.11	0.12	0.13	0.17	0.26	0.32	0.34	0.35	0.54	0.68	0.69	1
ɸ	0.23	0.22	0.22	0.21	0.17	0.16	0.13	0.11	0.08	0.07	0.03	0	0.03	0.04	0.05	0.06	0.07	0.09	0.1	0.11	0.15	0.24	0.3	0.33	0.34	0.53	0.67	0.69	1
f	0.25	0.24	0.24	0.23	0.19	0.18	0.15	0.13	0.1	0.09	0.05	0.03	0	0.01	0.03	0.04	0.05	0.06	0.08	0.09	0.12	0.22	0.28	0.31	0.32	0.52	0.67	0.68	1
v	0.26	0.25	0.25	0.24	0.2	0.19	0.16	0.14	0.11	0.1	0.06	0.04	0.01	0	0.02	0.03	0.04	0.05	0.06	0.08	0.11	0.21	0.28	0.3	0.31	0.51	0.66	0.68	1
θ	0.27	0.26	0.26	0.25	0.21	0.2	0.17	0.15	0.12	0.11	0.07	0.05	0.03	0.02	0	0.01	0.03	0.04	0.05	0.06	0.1	0.2	0.27	0.29	0.3	0.51	0.66	0.67	1
ð	0.28	0.27	0.26	0.26	0.22	0.21	0.18	0.16	0.14	0.13	0.08	0.06	0.04	0.03	0.01	0	0.01	0.03	0.04	0.05	0.09	0.19	0.26	0.28	0.3	0.5	0.65	0.67	1
s	0.29	0.28	0.27	0.27	0.23	0.22	0.19	0.17	0.15	0.14	0.1	0.07	0.05	0.04	0.03	0.01	0	0.01	0.03	0.04	0.08	0.18	0.25	0.27	0.29	0.49	0.65	0.66	1
z	0.3	0.29	0.28	0.28	0.24	0.23	0.2	0.18	0.16	0.15	0.11	0.09	0.06	0.05	0.04	0.03	0.01	0	0.01	0.03	0.07	0.17	0.24	0.26	0.28	0.49	0.64	0.66	1
ʃ	0.31	0.3	0.29	0.29	0.25	0.24	0.21	0.19	0.17	0.16	0.12	0.1	0.08	0.06	0.05	0.04	0.03	0.01	0	0.01	0.06	0.16	0.23	0.25	0.27	0.48	0.64	0.65	1
ʒ	0.32	0.31	0.3	0.3	0.26	0.25	0.22	0.2	0.18	0.17	0.13	0.11	0.09	0.08	0.06	0.05	0.04	0.03	0.01	0	0.04	0.15	0.22	0.24	0.26	0.47	0.64	0.65	1
ç	0.34	0.34	0.33	0.32	0.29	0.28	0.25	0.24	0.21	0.2	0.17	0.15	0.12	0.11	0.1	0.09	0.08	0.07	0.06	0.04	0	0.11	0.19	0.21	0.23	0.45	0.62	0.63	1
h	0.42	0.41	0.41	0.4	0.37	0.36	0.34	0.32	0.3	0.29	0.26	0.24	0.22	0.21	0.2	0.19	0.18	0.17	0.16	0.15	0.11	0	0.09	0.12	0.13	0.38	0.57	0.59	1
ɹ	0.46	0.46	0.45	0.45	0.42	0.41	0.39	0.38	0.36	0.35	0.32	0.3	0.28	0.28	0.27	0.26	0.25	0.24	0.23	0.22	0.19	0.09	0	0.04	0.06	0.33	0.53	0.55	1
j	0.48	0.48	0.47	0.47	0.44	0.43	0.41	0.4	0.38	0.37	0.34	0.33	0.31	0.3	0.29	0.28	0.27	0.26	0.25	0.24	0.21	0.12	0.04	0	0.03	0.3	0.52	0.54	1
ɰ	0.49	0.49	0.48	0.48	0.45	0.44	0.42	0.41	0.39	0.38	0.35	0.34	0.32	0.31	0.3	0.3	0.29	0.28	0.27	0.26	0.23	0.13	0.06	0.03	0	0.29	0.51	0.53	1
w	0.64	0.64	0.63	0.63	0.61	0.61	0.59	0.58	0.57	0.56	0.54	0.53	0.52	0.51	0.51	0.5	0.49	0.49	0.48	0.47	0.45	0.38	0.33	0.3	0.29	0	0.17	0.19	0.57
ɕ	0.75	0.75	0.75	0.74	0.73	0.73	0.72	0.71	0.7	0.7	0.68	0.67	0.67	0.66	0.66	0.65	0.65	0.64	0.64	0.64	0.62	0.57	0.53	0.52	0.51	0.17	0	0.01	0.33
ʑ	0.76	0.76	0.75	0.75	0.74	0.74	0.73	0.72	0.71	0.71	0.69	0.69	0.68	0.68	0.67	0.67	0.66	0.66	0.65	0.65	0.63	0.59	0.55	0.54	0.53	0.19	0.01	0	0.32
ɬ	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0.57	0.33	0.32	0

TABLE 4.4 Vowels distance table

	i	e	ɛ	a	ɑ	ɔ	o	u	ɯ	ɪ	ʊ	ə	æ
i	0	0.1	0.12	0.14	0.2	0.24	0.28	0.3	0.63	0.75	0.84	0.88	1
e	0.1	0	0.11	0.12	0.18	0.22	0.24	0.29	0.62	0.74	0.83	0.87	1
ɛ	0.12	0.11	0	0.1	0.16	0.17	0.23	0.26	0.6	0.73	0.83	0.87	1
a	0.14	0.12	0.1	0	0.11	0.17	0.2	0.23	0.58	0.72	0.82	0.86	1
ɑ	0.2	0.18	0.16	0.11	0	0.11	0.14	0.16	0.55	0.71	0.8	0.85	1
ɔ	0.24	0.22	0.17	0.17	0.11	0	0.14	0.15	0.53	0.69	0.79	0.84	1
o	0.28	0.24	0.23	0.2	0.14	0.14	0	0.13	0.51	0.68	0.78	0.83	1
u	0.3	0.29	0.26	0.23	0.16	0.15	0.13	0	0.47	0.66	0.76	0.82	1
ɯ	0.63	0.62	0.6	0.58	0.55	0.53	0.51	0.47	0	0.26	0.36	0.42	0.62
ɪ	0.75	0.74	0.73	0.72	0.71	0.69	0.68	0.66	0.26	0	0.16	0.22	0.35
ʊ	0.84	0.83	0.83	0.82	0.8	0.79	0.78	0.76	0.36	0.16	0	0.12	0.24
ə	0.88	0.87	0.87	0.86	0.85	0.84	0.83	0.82	0.42	0.22	0.12	0	0.19
æ	1	1	1	1	1	1	1	1	0.62	0.35	0.24	0.19	0

4.2. Experimental Design

For the experiment, we prepared txt format lists containing words and their corresponding IPA numbers for both English and Japanese datasets. Each line in the lists is structured as "word [consonant IPA numbers (Array 1)] [vowel IPA numbers (Array 2)] [all IPA numbers (Array 3)]." The English list consists of 8,325 words, while the Japanese list contains 16,719 words. Below, Fig. 4.1 presents a portion of the English word - IPA number list whereas Fig. 4.2 is for the Japanese word - IPA number list.

```
あんだけ [120, 104, 109] [304, 304, 302] [304, 120, 104, 304, 109, 302]
あんなに [120, 116, 116] [304, 304, 301] [304, 120, 116, 304, 116, 301]
あんべし [120, 102, 182] [304, 302, 301] [304, 120, 102, 302, 182, 301]
あんまし [120, 114, 182] [304, 304, 301] [304, 120, 114, 304, 182, 301]
あんま機 [120, 114, 109] [304, 304, 301] [304, 120, 114, 304, 109, 301]
あん肝 [120, 109, 114] [304, 301, 307] [304, 120, 109, 301, 114, 307]
いき値 [109, 103, 182] [301, 301, 301] [301, 109, 301, 103, 182, 301]
そ及 [132, 109, 153] [307, 316, 316] [132, 307, 109, 153, 316, 316]
そ求 [132, 109, 153] [307, 316, 316] [132, 307, 109, 153, 316, 316]
ぞくっと [133, 109, 103] [307, 316, 307] [133, 307, 109, 316, 103, 307]
```

Fig. 4.1. English word - IPA numbers list.

```
accion [103, 134, 116] [325, 301, 322] [325, 103, 134, 301, 322, 116]
acclaims [109, 114, 133] [322, 302, 319] [322, 109, 302, 319, 114, 133]
acclaimed [109, 114, 104] [322, 302, 319] [322, 109, 302, 319, 114, 104]
acclaims [109, 114, 133] [322, 302, 319] [322, 109, 302, 319, 114, 133]
account [109, 116, 103] [322, 304, 321] [322, 109, 304, 321, 116, 103]
accruals [109, 151, 133] [322, 308, 322] [322, 109, 151, 308, 322, 133]
domangue [104, 114, 119] [307, 321, 305] [104, 307, 321, 114, 305, 119]
domas [104, 114, 132] [307, 321, 322] [104, 307, 321, 114, 322, 132]
domecq [104, 114, 109] [307, 321, 303] [104, 307, 321, 114, 303, 109]
domek [104, 114, 109] [307, 321, 303] [104, 307, 321, 114, 303, 109]
```

Fig. 4.2. Japanese word - IPA numbers list.

In the program, when comparing IPA numbers, we use branching conditions referring to the corresponding arrays and the distance tables. To compare vowels, the program applied Array 1 and the "Consonants distance table" (TABLE 4.3), and to compare consonants, we use Array 2 and the "Vowels distance table" (TABLE 4.4) to calculate the distances. When comparing all IPA numbers, we check if the n-th IPA number in both Array 3 (for English and Japanese) is a consonant or a vowel. If they are both consonants, we refer to the "Consonants distance table" (TABLE 4.3), and if they are both vowels, we refer to the "Vowels distance table" (TABLE 4.4) to find the corresponding value. Otherwise, we return 1 as the maximum distance.

4.3. Results

By repeating the process in section 4.2, we calculated the overall distance between words as the sum of the distances between each pair of IPA numbers. As a result, we found the top 10 pairs of English-Japanese words with the closest distances between them. Fig. 4.3, Fig. 4.4, and Fig. 4.5 present the results for the three evaluation criteria: consonants, vowels, and all IPA numbers.

First, we compared the distances of individual consonants in each word. Fig. 4.3 shows the output of finding the top 10 pairs of words with the closest average distance between consonants in the list.

```
English abacus, Japanese: ＶＡＸ, Distance: 0.0
English abacus, Japanese: ばこそ, Distance: 0.0
English abacus, Japanese: バッカス, Distance: 0.0
English abacus, Japanese: バックス, Distance: 0.0
English abacus, Japanese: バックソー, Distance: 0.0
English abacus, Japanese: ボクサー, Distance: 0.0
English abacus, Japanese: ボックス, Distance: 0.0
English abacus, Japanese: 化かす, Distance: 0.0
English abacus, Japanese: 魅す, Distance: 0.0
English abacus, Japanese: 暈かす, Distance: 0.0
```

Fig. 4.3. 10 pairs of words with the highest consonant distances.

Similarly, we obtained the top 10 pairs of words with the closest average distance between vowels in the same list (shown in Fig. 4.4).

```
English bbc, Japanese: ＥＣＰ, Distance: 0.0
English bbc, Japanese: いき値, Distance: 0.0
English bbc, Japanese: いりちー, Distance: 0.0
English bbc, Japanese: きしり, Distance: 0.0
English bbc, Japanese: きりり, Distance: 0.0
English bbc, Japanese: ぎっしり, Distance: 0.0
English bbc, Japanese: ひひいん, Distance: 0.0
English bbc, Japanese: びしり, Distance: 0.0
English bbc, Japanese: びっしり, Distance: 0.0
English bbc, Japanese: びびり, Distance: 0.0
```

Fig. 4.4. 10 pairs of words with the highest vowel distances.

However, in both experiments, the results show only pairs with a distance of 0.0 at the top because there were more than 10 pairs of words holding completely identical elements. To verify the performance accuracy of the program and the distance tables, we calculated a total distance considering the order of consonants and vowels. Fig. 4.5 displays the top 10 pairs of words with the closest average distance between both consonants and vowels.

```
English minniti, Japanese: 耳ピ, Distance: 0.01
English ddt, Japanese: ビビビッ, Distance: 0.01
English ddt, Japanese: ピーピーピー, Distance: 0.01
English bbc, Japanese: びびり, Distance: 0.02
English bbc, Japanese: ビビリ, Distance: 0.02
English siddiqi, Japanese: 響, Distance: 0.02
English gdp, Japanese: ビビビッ, Distance: 0.02
English gdp, Japanese: ピーピーピー, Distance: 0.02
English gdp, Japanese: 忌引, Distance: 0.02
English gdp, Japanese: 忌引き, Distance: 0.02
```

Fig. 4.5. 10 pairs of words with the closest comprehensive distances.

5. Conclusion

5.1. Discussions

Through this experiment, we successfully identified pairs of words with similar pronunciations across different languages using the proposed algorithm. The most significant contribution of this experiment is that we assigned weights to the phonetic features, which allowed us to represent the distances between IPA symbols in the form of a distance table. Consequently, we could perform the objective evaluation as numerical values to measure the similarity between words, which was subjectively assessed in the linguistic research field. This provides a more robust and persuasive method for evaluating the phonetic characteristics, enhancing our understanding of cross-linguistic pronunciation patterns.

5.2. Limitation of study

We intended to conduct the human evaluation survey using the identified word pairs, as a comparative measure against the theoretically computed results. However, we could not create the questionnaire form due to time constraints. Additionally, this experiment was the initial stage of our research so we restricted the comparison set to words containing exactly three consonants and three vowels. To ensure the algorithm's broader applicability in the future, some modifications such as incorporating techniques (e.g. Dynamic Time Warping) are considerable to calculate distances for words of varying lengths.

5.3. Future research

We strongly believe that future research can consider the variety of evaluation methods for giving a solution to potential challenges in acquiring unfamiliar pronunciation in foreign languages. By comparing the phonetic aspects of languages, we can identify common areas of difficulty and provide insights into the specific features that learners should focus on to improve their pronunciation skills. Future research will enable us to explore a deeper understanding of the phonetic characteristics of various languages and provide a valuable resource for language learners, educators, and researchers.

References

[1]David Odden, Introducing Phonology, Cambridge University Press, 2005. (Online ISBN: 9780511808869, https://doi.org/10.1017/CBO9780511808869)

[2]O. Htun, S. Kodama, Y. Mikami. "Cross-language phonetic similarity measure on terms appeared in asian languages." *International Journal of Intelligent Information Processing*, vol. 2, no. 2, pp. 9-21, June 2011. (https://doi.org/10.4156/ijiip.vol2.issue2.2.)

[3]Jeffrey M. Silbert, "The World's First Computerized Criminal-Justice Information-Sharing System-The New York State Identification and Intelligence System (NYSIIS)." *Criminology*, vol. 8, no. 2, 107-128, 1970. (doi: 10.1111/j.1745-9125.1970.tb00734.x)

[4]Lawrence Philips, "Hanging on the Metaphone." *Computer Language*, vol. 7, no. 12, pp. 39-44, Dec. 1990.

[5]Pimpalkhute, Pranoti, et al. "Phonetic spelling filter for keyword selection in drug mention mining from social media." *AMIA Summits on Translational Science Proceedings* 2014, pp. 90-95, 2014.

[6]Lawrence Philips, "The double metaphone search algorithm." *C/C++ users journalI,* vol. 18, no. 6, pp. 38-43, June 2000.

[7]Kevin Atkinson, "Gnu aspell 0.60. 4." 2006.

[8]International Phonetic Association, "The international phonetic alphabet and the IPA chart." March 2015.

[9]Verma, Dhruv, and S. N. Muralikrishna. "Semantic similarity between short paragraphs using Deep Learning." *2020 IEEE International Conference on Electronics, Computing and Communication Technologies (CONECCT)*, IEEE, Jul 2020, pp. 1-5. (doi: 10.1109/CONECCT50063.2020.9198445)