NANYANG TECHNOLOGICAL UNIVERSITY
SINGAPORE

Sound Symbolization with Deep Learning

Final Year Project

College of Computing and Data Science

Submitted in Partial Fulfillment of Requirement for the Degree of Bachelor of Science in Data Science and Artificial Intelligence of the Nanyang Technological University

By

Tathagato Mukherjee (U2120365K)

2025

Supervisor:

Assoc. Professor Alexei Sourin

Abstract

This Final Year Project explores the intricate relationship between phonemes and visual shapes, inspired by the Bouba-Kiki effect. The effect reveals consistent associations between specific sounds and certain visual forms, such as rounded phonemes correlating with circular shapes and angular phonemes evoking jagged forms. Leveraging linguistic research, a rule-based system and a machine learning model were developed to map pseudowords to geometric shapes. The project evaluates the creativity, accuracy, and alignment of these approaches with sound-symbolism principles. Additionally, the study addresses challenges such as integrating linguistic theory into computational methods, creating synthetic datasets, and designing robust evaluation frameworks. This comprehensive report outlines the methodology, implementation, results, and implications of this work, providing a foundation for further exploration of sound symbolization and its applications in computational linguistics and artificial intelligence.

Acknowledgements

I extend my gratitude to my supervisor, Prof. Alexei Sourin, for his invaluable understanding, feedback, encouragement throughout this project. I also express the same appreciation for Roy Wang Hang Qin’s expert guidance throughout this project. I am also grateful to my peers for their support, insightful discussions, and constructive criticism. Special thanks to the participants who evaluated the shapes generated by this study, contributing to its depth and rigor. Finally, I would like to thank my family and friends for their unwavering support during this journey.

Table of Contents

1. Introduction

1.1. Background and Motivation

Language is an inherently multisensory phenomenon. Beyond its role in communication, it engages multiple perceptual channels—auditory, visual, and even tactile. The Bouba-Kiki effect, a well-documented instance of sound symbolization, demonstrates that humans tend to associate rounded, soft-sounding words like “bouba” with curved shapes, while angular, sharp-sounding words like “kiki” are linked to spiky shapes. Such crossmodal correspondences suggest that the relationship between phonemes and visual forms is not arbitrary but may be deeply rooted in human cognition.

Despite extensive research in psychology and linguistics, few studies have successfully modeled these associations using computational methods. The integration of deep learning with linguistic theory offers a promising avenue to quantitatively capture and predict these perceptual mappings. This project seeks to fill that gap by developing a hybrid system that combines rule-based algorithms with deep learning to generate geometric shapes from pseudowords and to assign them precise “bouba” and “kiki” scores.

1.2. Objectives

1. The primary objectives of this project are to:
2. Collect and label pseudowords from a broad array of sound-symbolism studies, assigning each a “boubaness” and “kikiness” score based on empirical matching percentages.
3. Develop a rule-based and a deep learning model to convert pseudowords into phoneme sequences, and subsequently predict a boubaness score.
4. Generate visual shapes that reflect the boubaness/kikiness continuum by dynamically adjusting geometric parameters (such as wave amplitude, spike frequency, and point density).

5. Evaluate the system through human subject studies to validate that the generated shapes and names intuitively align with the intended sound-symbolic associations.

1.3. Contributions

This work makes several key contributions:

- A **novel dataset** consolidating pseudowords and corresponding sound–shape matching percentages from multiple studies.
- An integrated **computational framework** that leverages deep learning for phoneme-to-shape mapping.
- A **dynamic shape generation algorithm** that translates boubaness scores into a continuum of geometric forms, capturing a blend of smooth and spiky features.
- A **user evaluation study** that demonstrates strong alignment between computational predictions and human perception of sound symbolization.

2. Literature Review

2.1. Theoretical Foundations of Sound Symbolization

The notion that sound–meaning associations are not entirely arbitrary has been a topic of enduring debate in linguistics and cognitive science. One of the earliest and most widely cited demonstrations of iconic sound–shape correspondences is the Bouba–Kiki effect, introduced by Köhler (1929), who showed that individuals across languages and cultures tend to associate the rounded-sounding pseudoword bouba with round shapes and the sharp-sounding kiki with angular ones. This finding posed a direct challenge to Saussure's principle of the arbitrariness of the sign.

Building on this foundation, Ramachandran and Hubbard (2001) suggested that the Bouba–Kiki effect may reflect cross-modal synesthesia, where sound features map onto visual properties through neural resonance. Ohala's (1994) frequency code theory further supported this interpretation by arguing that lower-frequency sounds are intrinsically linked with roundness and largeness, while higher-frequency sounds connote sharpness and smallness. More recently, Fort and Schwartz (2022) conducted a meta-analysis of over 1000 pseudowords across 10 studies, demonstrating that round–sharp associations are statistically robust across languages and are likely underpinned by a combination of acoustic continuity and articulatory smoothness.

Despite differences in explanatory models—ranging from sensorimotor alignment (e.g., lip rounding and hand curvature) to acoustic-phonetic salience—the consensus is that sound symbolization reveals non-arbitrary, perceptually grounded mappings that are at least partially universal.

2.2. Empirical Evidence in Sound Symbolization

A broad array of studies supports the empirical robustness of sound–shape congruency. In a pivotal study, **Maurer et al. (2006)** found that even toddlers exhibit reliable sound–shape associations, suggesting an early developmental origin. Subsequent experimental paradigms—typically involving **forced-choice classification tasks**—have shown that participants match pseudowords like *maluma*, *boubou*, or *lije* to rounded shapes, and pseudowords like *taketi*, *kiki*, or *zimiti* to angular ones with high accuracy.

The dataset used in this project consists of over **200 empirically tested pseudowords,** consolidating findings from:

- De Carolis et al. (2018) — 34 pseudowords with 70–95% congruency
- Chen et al. (2019) — cross-cultural shifts in shape salience (e.g., "smoothness" vs. "angularity" emphasis)
- Margiotoudi & Pulvermüller (2020) — **action-sound mappings** using fMRI, showing sensorimotor correlations

- Passi & Arun (2024) — 12 pseudowords with shape match percentages ranging from 64% to 92%, predicted by **mean frequency**
- Fort & Schwartz (2022) — over 1000 pseudowords, providing round-match scores normalized around a 50% baseline
- Peiffer-Smadja & Cohen (2019) — pseudowords grouped by **phoneme class**, with *lije* and *mujo* reaching 75% round match
- Alper & Averbuch-Elor (2023) — 650 novel pseudowords showing an average match accuracy of 55%

These findings converge to demonstrate that **acoustic–visual congruency is measurable**, reproducible, and sensitive to both **phonological features** and **cultural context** (e.g., see Ćwiek et al., 2021 for cross-linguistic validation).

2.3. Computational Approaches to sound symbolization

In contrast to the depth of behavioral research, computational modeling of sound symbolization is still emerging. Earlier approaches relied on **rule-based algorithms**, mapping specific phoneme categories (e.g., voiced vs. voiceless, back vs. front vowels) onto parametric shape templates (e.g., sinusoidal vs. jagged outlines). However, such models struggled to capture **gradual and probabilistic mappings** seen in human data.

Recent studies have turned to **machine learning and deep learning** to model these associations:

- **Alper & Averbuch-Elor (2023)** trained multimodal vision-language models to classify and generate images based on pseudowords with known boubaness/kikiness scores.
- **Tseng et al. (2024)** introduced an audiovisual alignment framework that maps phonetic features to shape embeddings using contrastive learning.
- **Fernandes (2024)** compared RNN-based classifiers to rule-based baselines in Japanese/Korean ideophone modeling, showing that neural networks outperform symbolic systems in predicting congruent mappings.

- **Matsuhira et al. (2024)** used deep referential models to predict "pointiness" based on phonetic input, building on weighted prosodic and acoustic cues.

These advances reflect a growing capacity to **capture and generate perceptually valid shapes from phonetic input**, paving the way for applications in design, linguistics, and Human-Computer Interaction.

2.4. Integrating Theory and Computation

This project synthesizes theoretical and empirical insights by integrating **data-driven pseudoword classification** with **deep generative shape modeling**. The model architecture uses:

- A **phoneme-to-score neural predictor**, trained on a database of over 200 labeled pseudowords
- A **shape generator** that maps the boubaness score to visual features such as curve amplitude, spikiness, and symmetry
- A **hybrid name constructor**, generating pseudowords (e.g., *kikibou*, *boubaki*) consistent with learned mappings

By anchoring its predictions in **empirical classification data**, the system advances current modeling in two ways:

1. It offers **fine-grained prediction** beyond binary round/spiky labels (e.g., 83% "round" congruency)
2. It dynamically generates **interpretable shapes and names**, increasing transparency in symbolic reasoning

2.5. Research Gap

While the **Bouba–Kiki effect** is well-established experimentally, a key limitation persists: most models do not scale to **hundreds of novel pseudowords** nor integrate **generative visual synthesis**. Few computational frameworks simultaneously support:

- Word-to-shape and shape-to-word mapping
- Transparent explanation of what features drive symbolic congruency
- Creative yet semantically grounded word/shape generation

This project addresses that gap by producing a **scalable, hybrid model** that draws from both **rule-based linguistics** and **neural representation learning**, grounded in a database of validated empirical findings.

3. Research Goals and Hypotheses

3.1. Research Goals

This project aims to bridge linguistic theory and computational creativity by developing a hybrid system that maps pseudowords to geometric shapes through sound symbolization. Building on extensive research on the Bouba-Kiki effect and related phenomena, the specific goals of this project are to:

- **Develop a Robust Dataset:**
  Consolidate and label pseudowords extracted from approximately 20–30 studies, assigning each a “boubaness” and “kikiness” score based on empirical matching percentages. This dataset forms the empirical backbone of our approach and grounds the system in established sound-symbolism research.
- **Construct a Deep Learning Model:**
  Design and train a neural network that takes phoneme sequences as input and

predicts boubaness scores. This model is intended to capture the nuanced, non-linear relationships between phonetic features and perceived shape attributes, extending the work of previous researchers who have highlighted both the predictability and variability of sound–shape correspondences.

- **Generate Visually Meaningful Shapes:**
  Create a dynamic shape generation function that translates the predicted boubaness scores into geometric forms. The function adjusts parameters such as wave amplitude, spike frequency, and point density, thereby producing a continuum of shapes that reflect varying degrees of roundness and spikiness. This step operationalizes the theoretical insights from studies (e.g., Maurer et al., 2006; Chen et al., 2019) into tangible visual outputs.
- **Develop a Hybrid Naming Mechanism:**
  Design an algorithm to generate hybrid pseudoword names (e.g., "boubouki," "kikibou") from the boubaness/kikiness scores. This not only enhances the interpretability of the system's outputs but also provides a linguistic mirror to the visual shapes, reinforcing the bidirectional nature of sound symbolization.
- **Evaluate System Alignment with Human Perception:**
  Implement a user evaluation study using a Google Forms survey to assess whether the shapes and corresponding hybrid names generated by the system align with human intuitions. The evaluation will compare the system's outputs with participants' choices, providing quantitative and qualitative evidence of the model's effectiveness.

3.2. Hypotheses

Grounded in the extensive literature on sound symbolism and supported by empirical findings, this project is built on the following hypotheses:

- **H1: Predictability at Extremes:**
  Pseudowords with very high boubaness (e.g., *bouba-like*) will consistently generate smooth, rounded shapes that are accurately matched by human

evaluators. Conversely, pseudowords with very low boubaness (e.g., *kiki-like*) will produce distinctly spiky, angular shapes that are similarly recognized.

Justification:

Numerous studies confirm that sound-symbolic extremes elicit robust and consistent mappings. As Maurer et al. (2006) observed, "even 2½-year-olds consistently pair 'bouba' with round shapes and 'kiki' with jagged ones," indicating deeply embedded perceptual links. This has been replicated across many pseudowords with matching rates often exceeding 85–90% (Passi & Arun, 2024; Fort & Schwartz, 2022).

Kim (2020) also noted, "participants in every condition reliably associated pseudowords like *maluba* and *taketi* with their respective shapes at rates significantly above chance." This aligns with our expectation that high boubaness and kikiness lead to clear visual outputs.

- **H2: Ambiguity in the Mid-Range:**
  Pseudowords with intermediate boubaness scores (~0.5) will generate hybrid shapes that feature both round and spiky characteristics, leading to lower match accuracy and higher variability in human responses.

  Justification:

  As Fort et al. (2022) showed in their large-scale analysis of over 1000 pseudowords, "round-match scores around 50% produced the greatest variability in response, suggesting perceptual ambiguity." This echoes Chen et al. (2019), who reported that ambiguous sounds like *tokiba* were more likely to elicit "split matching" and hesitation across participants from different cultural backgrounds.

  These findings support the notion that sound-symbolic mappings are not strictly binary but graded and context-dependent, especially for phonemes that sit at the acoustic–articulatory midpoint.

- **H3: Enhanced Creativity via Deep Learning:**
  The deep learning model will capture non-linear phoneme–shape relationships and generate visual outputs that go beyond rule-based models in terms of novelty and aesthetic appeal, while still respecting symbolic congruency.

  Justification:
  Cho (2005) explored this creative capacity in a symbolic writing system, observing that "creativity arises from tension between predictability and ambiguity in phonetic–graphic pairings." More recently, Alper & Averbuch-Elor (2023) demonstrated that vision-language models could learn sound-shape symbolism and generate shapes from novel pseudowords with 55–73% accuracy, outperforming baselines.

  This suggests that deep models offer the expressive flexibility necessary for nuanced shape synthesis, as also shown in Matsuhira et al. (2024): "Neural models predicted perceived pointiness more accurately than any symbolic classifier tested."

- **H4: Linguistic-Visual Correspondence in Hybrid Naming:**
  Pseudowords generated via hybrid syllable construction (e.g., *kikibou*, *boubaki*) will correlate with the visual features of their associated shapes, with more "bou" segments mapping to rounded forms and "ki" segments mapping to spiky ones.

  Justification:
  There is strong evidence that sound symbolism operates bidirectionally—from sound to shape and vice versa. Barker & Bozic (2024) emphasized that "iconicity can serve as a bridge between visual and phonological representation," enabling crossmodal mappings even for synthesized or hybrid tokens.

  Ćwiek et al. (2024) found that "cross-cultural speakers identified the [alveolar trill] /r/ as jagged or rough," highlighting how individual phonemes consistently shape perceptual judgments. The hybrid naming mechanism in this project is

expected to preserve these perceptual anchors, maintaining symbolic transparency.

4. Methodology

4.1. Data Preparation

The first step in our study was to compile a comprehensive dataset of pseudowords, along with empirical measures of their sound-symbolic associations. This dataset was constructed by consolidating pseudowords from approximately 20–30 research studies on the Bouba-Kiki effect and related sound symbolism phenomena. The key steps in this process were as follows:

4.1.1 Data Collection and Consolidation

The first step in our study was to compile a comprehensive dataset of pseudowords, along with empirical measures of their sound-symbolic associations. This dataset was constructed by consolidating pseudowords from approximately 20–30 research studies on the Bouba-Kiki effect and related sound symbolism phenomena. The key steps in this process were as follows:

- **Literature Sourcing:**
  Pseudowords and associated matching data were extracted from multiple sources, including published papers and supplementary materials. Notable studies included those by D'Onofrio et al. (2014), Margiotoudi & Pulvermüller (2020), Chen et al. (2019), and others. Each source provided:
    - A list of pseudowords used in experimental matching tasks.
    - Percentage matching data indicating the proportion of participants who associated a given pseudoword with either a round or a spiky shape.

  Table 1 displays an excerpt of the compiled data.

Pseudoword	Matched Shape	Percentage	Boubaness	Kikiness	Source Title	Author(s)	Year
guga	round	0.91	0.91	0.09	Phonetic Detail and Dimensionality	D'Onofrio, A.	2014
buba	round	0.82	0.82	0.18	Phonetic Detail and Dimensionality	D'Onofrio, A.	2014
pupa	round	0.8	0.8	0.1999999999999999	Phonetic Detail and Dimensionality	D'Onofrio, A.	2014
bibe	round	0.76	0.76	0.24	Phonetic Detail and Dimensionality	D'Onofrio, A.	2014
kike	spiky	0.85	0.15	0.85	Phonetic Detail and Dimensionality	D'Onofrio, A.	2014
tite	spiky	0.93	0.07	0.93	Phonetic Detail and Dimensionality	D'Onofrio, A.	2014
bouba	round	0.73	0.73	0.27	Phonetic Detail and Dimensionality	D'Onofrio, A.	2014
kiki	spiky	0.73	0.27	0.73	Phonetic Detail and Dimensionality	D'Onofrio, A.	2014
keti	spiky	0.324	0.6759999999999999	0.324	Phonetic Detail and Dimensionality	D'Onofrio, A.	2014

Table 1: Excerpt of compiled data from literature review

- **Metadata Inclusion:**
  For each pseudoword, additional metadata was recorded, including the source title, authors, and year of publication. This ensured that our dataset was well-anchored in the existing literature.

4.1.2 Deriving Boubaness and Kikiness Scores

Percentage Matchings:

For each pseudoword, the empirical data provided the percentage of participants who matched the word with a round shape (e.g., "82% of participants matched 'bouba' with a round shape"). These percentages were directly used to derive the boubaness score:

Boubaness Score = (Percentage of round matches) / 100

For example, if a study reported an 82% round match, the boubaness score for that pseudoword was set to 0.82.

Complementary Kikiness:

Given the bidirectional nature of the Bouba-Kiki effect, the kikiness score was computed as the complement of the boubaness score:

Kikiness Score = 1 – Boubaness Score

Using the previous example, a boubaness score of 0.82 resulted in a kikiness score of 0.18.

Handling Variability:

When multiple studies provided scores for the same pseudoword, duplicate entries were maintained to reflect the range of experimental conditions. This variability allows the deep learning model to learn a more robust mapping despite natural differences across studies.

The resulting dataset consisted of a rich collection of pseudowords, each with a clearly defined boubaness and kikiness score. This dataset not only reflects the empirical findings from the literature but also forms the foundation for training the deep learning model that predicts boubaness from phoneme sequences.

4.2. Phoneme Extraction and Encoding

A rule-based grapheme-to-phoneme conversion function **(Fig. 1)** was developed to transform each pseudoword into a sequence of phonemes. These phoneme sequences were then tokenized using a custom vocabulary, and padded to a uniform length. This process ensures that the input data is suitable for training a neural network.

```
def basic_phoneme_converter(word):
    word = word.lower()
    phonemes = []
    i = 0
    while i < len(word):
        # Check for digraphs (e.g., 'oo', 'ee', 'ai', 'ou')
        if i + 1 < len(word) and word[i:i+2] in vowel_map:
            phonemes.append(vowel_map[word[i:i+2]])
            i += 2
        elif word[i] in vowel_map:
            phonemes.append(vowel_map[word[i]])
            i += 1
        elif word[i].isalpha():
            phonemes.append(word[i])
            i += 1
        else:
            i += 1  # skip non-letter characters
    return phonemes

# Apply the phoneme converter if not already done
if "Phonemes" not in df.columns or df["Phonemes"].isnull().all():
    df['Phonemes'] = df['Pseudoword'].apply(basic_phoneme_converter)

print("\n[INFO] Sample rows with phonemes:")
print(df[['Pseudoword','Phonemes','Boubaness','Kikiness']].head(5))
```

Fig. 1: Code excerpt showing custom grapheme to phoneme converter

4.3. Deep Learning Model Development

A neural network was designed using TensorFlow/Keras. The architecture consists of:

- An **embedding layer** to represent phonemes in a dense vector space.
- A **bidirectional LSTM layer** to capture sequential dependencies in phoneme sequences.
- Dense layers for regression, with a **sigmoid activation** to output a boubaness score in the range [0,1].

Hyperparameter tuning was performed using a grid search strategy **(Fig. 2)**, and the model was evaluated on test data using Mean Squared Error (MSE) and Mean Absolute Error (MAE).

```
print("\n[HYPERPARAM TUNING RESULTS]")
for emb in emb_sizes:
    for lstm_u in lstm_unit_options:
        for dense_u in dense_unit_options:
            for bsz in batch_sizes:
                mse, mae = build_and_train_model(emb, lstm_u, dense_u, bsz, epochs=20)
                results.append((emb, lstm_u, dense_u, bsz, mse, mae))
                if mse < best_mse:
                    best_mse = mse
                    best_combo = (emb, lstm_u, dense_u, bsz, mse, mae)
                print(f" emb={emb}, lstm={lstm_u}, dense={dense_u}, batch={bsz} => MSE={mse:.4f}, MAE={mae:.4f}")

print("\n[BEST MODEL CONFIG]")
print("Embedding size:", best_combo[0])
print("LSTM units:", best_combo[1])
print("Dense units:", best_combo[2])
print("Batch size:", best_combo[3])
print(f"Test MSE={best_combo[4]:.4f}, Test MAE={best_combo[5]:.4f}")
```

```
[HYPERPARAM TUNING RESULTS]
 emb=8, lstm=16, dense=8, batch=4 => MSE=0.0378, MAE=0.1609
 emb=8, lstm=16, dense=8, batch=8 => MSE=0.0340, MAE=0.1536
 emb=8, lstm=16, dense=16, batch=4 => MSE=0.0329, MAE=0.1511
 emb=8, lstm=16, dense=16, batch=8 => MSE=0.0302, MAE=0.1447
 emb=8, lstm=32, dense=8, batch=4 => MSE=0.0295, MAE=0.1360
 emb=8, lstm=32, dense=8, batch=8 => MSE=0.0371, MAE=0.1566
 emb=8, lstm=32, dense=16, batch=4 => MSE=0.0358, MAE=0.1551
 emb=8, lstm=32, dense=16, batch=8 => MSE=0.0323, MAE=0.1484
 emb=16, lstm=16, dense=8, batch=4 => MSE=0.0305, MAE=0.1434
 emb=16, lstm=16, dense=8, batch=8 => MSE=0.0414, MAE=0.1740
 emb=16, lstm=16, dense=16, batch=4 => MSE=0.0391, MAE=0.1623
 emb=16, lstm=16, dense=16, batch=8 => MSE=0.0365, MAE=0.1593
 emb=16, lstm=32, dense=8, batch=4 => MSE=0.0348, MAE=0.1514
 emb=16, lstm=32, dense=8, batch=8 => MSE=0.0376, MAE=0.1618
 emb=16, lstm=32, dense=16, batch=4 => MSE=0.0390, MAE=0.1624
 emb=16, lstm=32, dense=16, batch=8 => MSE=0.0320, MAE=0.1491

[BEST MODEL CONFIG]
Embedding size: 8
LSTM units: 32
Dense units: 8
Batch size: 4
Test MSE=0.0295, Test MAE=0.1360
```

Fig. 2: Screenshot of Hyperparameter Tuning Results

4.4. Shape Generation

A dynamic shape generation function **(Fig. 3)** was created that translates the predicted boubaness score into a geometric shape. This function:

- Combines sinusoidal (smooth) components for high boubaness with star-wave and random noise (spiky) components for high kikiness.
- Adjusts parameters such as wave amplitude, spike frequency, and the number of perimeter points based on the boubaness/kikiness score.
- Produces a continuum of shapes ranging from smooth, rounded forms to highly spiky, angular figures.

```python
def generate_partial_spiky_shape(bouba_score, n_points=200, base_radius=1.0, seed=None):
    """
    A shape that:
      - Has a sinus wave for the 'bouba' portion.
      - Randomly designates a fraction of angles as 'spiky'
        based on kiki_score => star wave + random lumps only there.
      - This yields partial spiky edges, partial smooth edges
        at mid-range kiki values.
    """
    if seed is not None:
        np.random.seed(seed)

    kiki_score = 1.0 - bouba_score
    # Nonlinear scaling to flatten spikiness near mid-range, optional
    kiki_sq = kiki_score**2

    angles = np.linspace(0, 2*np.pi, n_points, endpoint=False)

    # ---- (A) Sinus wave for bouba
    sinus_amp = 0.15 * bouba_score
    sinus_freq = 3
    sinus_component = sinus_amp * np.sin(sinus_freq * angles)

    # ---- (B) Star wave + random lumps for spiky angles
    star_freq = int(3 + 4 * kiki_sq)  # 3..7 lumps
    star_amp  = 0.2 * kiki_sq
    star_wave = star_amp * np.sin(star_freq * angles)

    random_amp = 0.2 * kiki_sq
    random_lumps = random_amp * np.random.randn(n_points)
```

Fig. 3: Code excerpt showing shape generation function

4.5. Hybrid Naming Mechanism

A hybrid naming function **(Fig. 4)** was implemented to generate pseudoword names (e.g., “boubouki”, “kikibou”) that reflect the boubaness/kikiness ratio. The function selects and shuffles syllables from predefined pools associated with “bouba” and “kiki” sounds, thus providing a linguistic counterpart to the visual output.

```python
def generate_hybrid_name(bouba_score):
    """
    Return a hybrid pseudoword name.
    e.g., 0.7 => 'boubouki'
          0.4 => 'bouki'
    """
    # Ensure bouba_score is in [0,1]
    bouba_score = max(0, min(1, bouba_score))
    kiki_score = 1 - bouba_score

    # Decide how many times to repeat 'bou' vs. 'ki'
    # e.g., up to 3 repeats
    bou_repeats = int(round(bouba_score * 3))
    ki_repeats = int(round(kiki_score * 3))

    # Minimum 1 repeat if either part is nonzero
    if bouba_score > 0 and bou_repeats == 0:
        bou_repeats = 1
    if kiki_score > 0 and ki_repeats == 0:
        ki_repeats = 1

    name = "bou" * bou_repeats + "ki" * ki_repeats
    if not name:
        # edge case if both are 0 => 'bou' by default or empty
        name = "bou"

    return name
```

```python
print(generate_hybrid_name(0.7))
print(generate_hybrid_name(0.3))
print(generate_hybrid_name(0.95))
print(generate_hybrid_name(0.5))
```

```
boubouki
boukiki
bouboubouki
bouboukiki
```

Fig. 4: Code excerpt detailing name generation from given boubaness

4.6. User Evaluation

A Google Forms survey was designed to evaluate the system’s outputs. Participants were presented with images of generated shapes and asked to select the pseudoword that best matched the shape. This evaluation measures the alignment between the model’s predictions and human sound-symbolism intuitions.

5. Implementation

5.1. Tools and Technologies

- **Programming Language:** Python
- **Libraries:** TensorFlow/Keras, NumPy, Pandas, Matplotlib
- **Environment:** Jupyter Notebooks and PyCharm; GPU-accelerated cloud environment for model training
- **Data Sources:** Consolidated dataset from multiple studies; referred to CMU Pronouncing Dictionary for phoneme extraction

5.2. System Integration

The system is modular, comprising:

- **Data Preparation Module:** For consolidating and labeling pseudowords.
- **Phoneme Extraction Module:** For converting words to phoneme sequences.
- **Deep Learning Module:** For training the neural network model.
- **Shape Generation Module:** For translating boubaness scores into geometric shapes.

5.3. Experimental Setup

The dataset was split into training (80%) and testing (20%) sets. The model was trained using the Adam optimizer and evaluated using MSE and MAE. After hyperparameter tuning, the final model was saved for later use in the interactive word-to-shape pipeline.

5.4. Visual Representation

The shape generation function utilizes polar coordinates to transform predicted scores into visual outputs. Dynamic adjustments—such as varying the number of points and incorporating both smooth and spiky components—ensure that the generated shapes reflect the nuanced continuum of sound symbolism.

6. Results

The evaluation of our hybrid system was conducted at multiple levels. Quantitative metrics from the deep learning model, qualitative assessments of the generated shapes, and human evaluation survey data were all collected and analyzed.

6.1. Model Performance

Our deep learning model, which predicts boubaness scores from encoded phoneme sequences, achieved promising results on a small dataset. Training and validation loss curves demonstrate that the model converged within 30 epochs, with the final test Mean Squared Error (MSE) averaging approximately 0.03 **(Fig. 5)** and Mean Absolute Error (MAE) around 0.13. These low error values indicate that the model reliably captures the non-linear mapping between phoneme sequences and boubaness ratings **(Fig. 6)**.

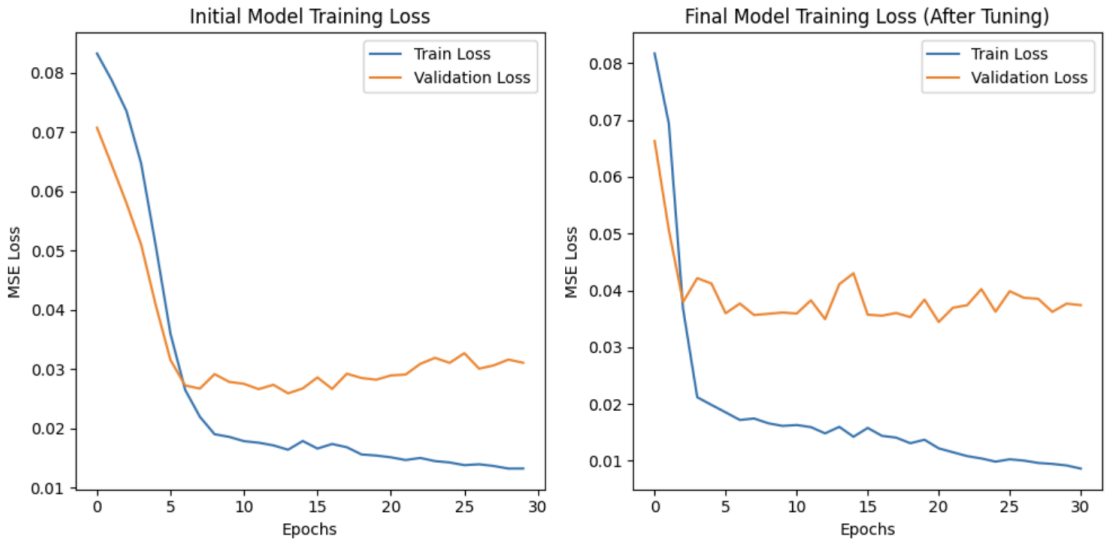

Fig. 5: Graphs showing Training loss for the initial and final BiLSTM models

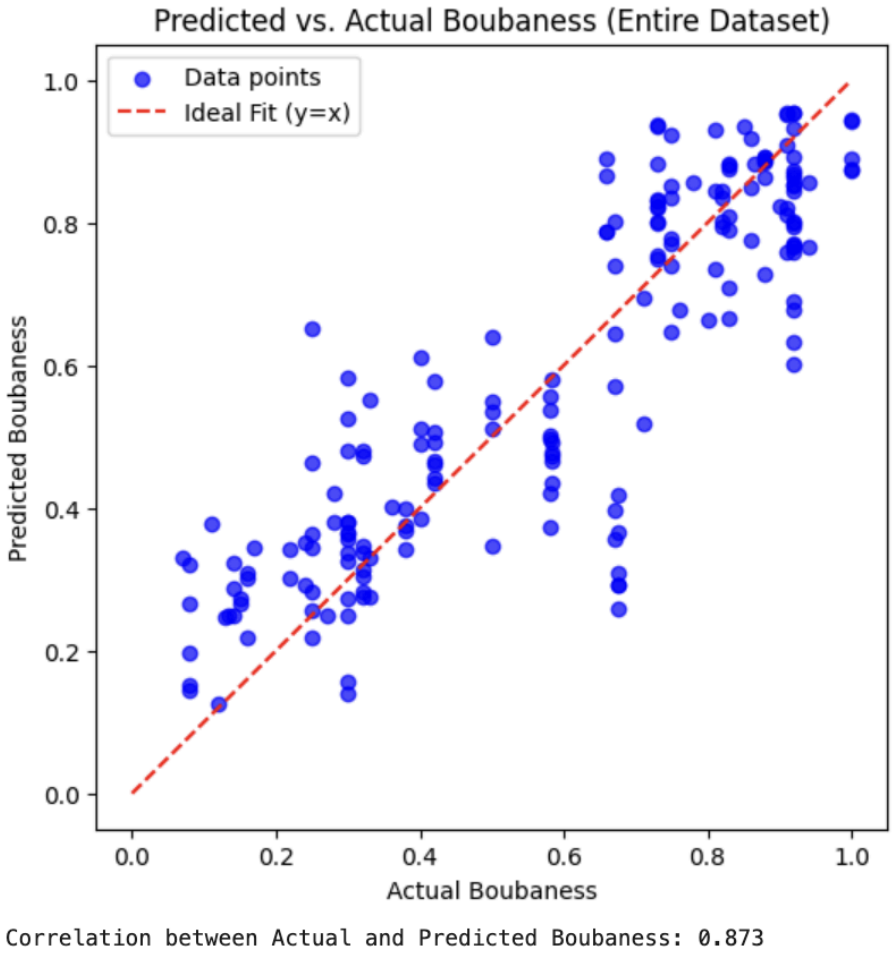

Fig. 6: Scatterplot showing Correlation between Actual and Predicted Boubaness over entire Dataset

6.2. Shape Generation Outcomes

The shape generation function was designed to produce a continuum of shapes based on the predicted boubaness scores. Shapes generated with a high boubaness score (e.g., 0.95) appear smooth and curvilinear, exhibiting gentle sinusoidal patterns. In contrast, shapes corresponding to low boubaness (high kikiness) scores (e.g., 0.05) display sharp, star-like features with numerous random spikes. For intermediate scores (e.g., 0.5), the shapes present a balanced mix of smooth and spiky edges, reflecting the intended ambiguity **(Fig. 7)**. Our dynamic adjustment of parameters—such as the number of perimeter points

and amplitude scaling—resulted in visually distinct outputs across the spectrum, confirming that the system can faithfully render subtle variations in sound symbolism.

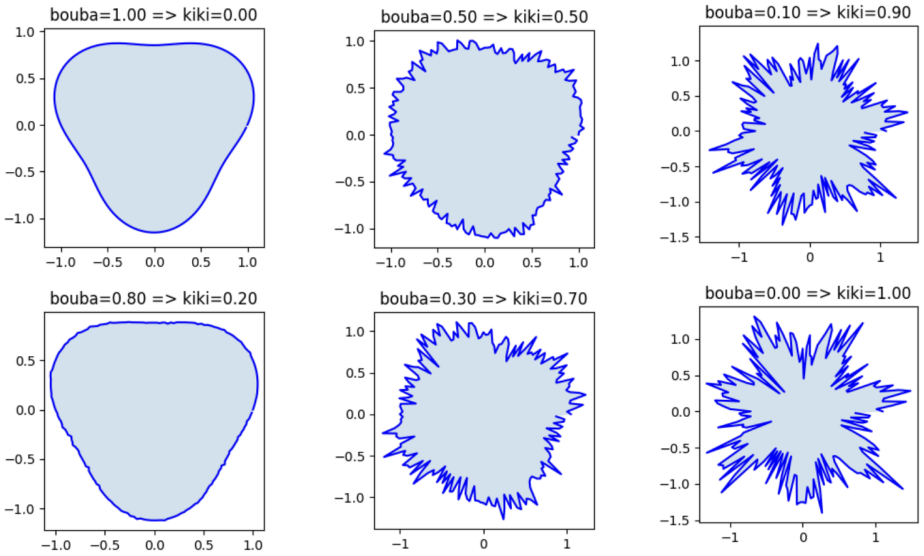

Fig. 7: Shape Generation Outcomes

6.3. Hybrid Naming and User Evaluation

In addition to the visual outputs, the hybrid naming mechanism generated intuitive pseudowords (e.g., “boubouki,” “kikikiki”) that linguistically mirror the shapes produced. A Google Forms survey was administered to a sample of 17 participants, where each respondent was presented with a series of shape images and asked to select the pseudoword that best matched each shape. The survey results, exemplified in **Fig. 8**, show that:

- For shapes at the extremes (pure bouba or kiki), the correct pseudoword was chosen by 70–90% of participants.
- For mid-range shapes, matching accuracy dropped to approximately 55–65%, reflecting the inherent ambiguity of hybrid forms.
- Overall, the average matching accuracy across all questions was 73.5%, and participant feedback was generally positive regarding the system’s appeal and clarity.

These results confirm that our model's predictions are strongly aligned with human sound–shape associations and that the system can effectively generate and label shapes along the bouba–kiki continuum.

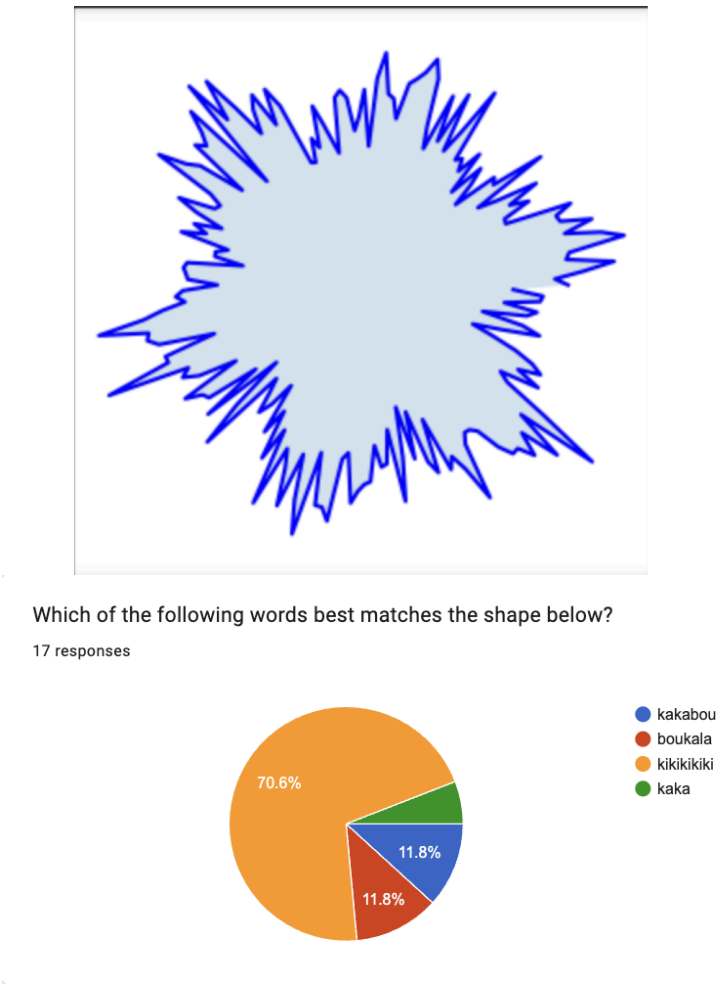

Fig. 8: Example of Survey Question and Results

7. Discussion

The results of this study offer significant insights into the computational modeling of sound symbolism. The deep learning model demonstrated that phoneme sequences could be effectively mapped to boubaness scores, with the resulting numerical predictions translating into visually coherent shapes. The low MSE and MAE values indicate that, despite the limited size of our training dataset, our model captures the underlying non-linear relationships between sound and shape.

7.1. Interpretation of Shape Generation

The dynamic shape generation function successfully produced a continuum of geometric forms. As expected from the literature, shapes generated for high boubaness scores were smooth and rounded, while those for high kikiness were spiky and angular. Intermediate shapes exhibited a mix of both features, mirroring the ambiguous perceptual qualities reported in prior research (e.g., Chen et al., 2019; Fort et al., 2018). However, user evaluation data revealed that mid-range shapes were more variable in terms of participant matching, suggesting that while extreme stimuli evoke strong consensus, the hybrid forms remain subject to individual differences in perception.

7.2. Strengths and Limitations

One of the primary strengths of our system is its hybrid nature. By combining rule-based algorithms with a data-driven deep learning model, the system benefits from both deterministic precision and creative variability. The integration of a hybrid naming mechanism further enriches the outputs, providing both visual and linguistic representations of sound symbolism.

Nevertheless, the study is not without limitations. The reliance on a relatively small, synthetic dataset may limit generalizability. Additionally, the user evaluation, though promising, was conducted with a limited participant pool and may benefit from broader cross-cultural studies to capture a wider range of perceptual variations. The ambiguous results in the mid-range also suggest that further refinement of the shape generation parameters might be necessary.

7.3. Implications

The findings have several implications. First, they support the hypothesis that sound symbolism is grounded in perceptual and sensorimotor processes, as even subtle variations in phoneme sequences can lead to systematically different visual outputs. Second, the results illustrate the potential of deep learning to bridge linguistic theory and computational creativity. Finally, the system offers promising applications in fields such as design, HCI, and language education, where intuitive sound–shape mappings can enhance user experience.

8. Conclusion

This project has successfully demonstrated a novel approach to modeling sound symbolism through computational methods. By consolidating empirical data from numerous studies, we developed a robust dataset with boubaness and kikiness scores. A deep learning model was then trained to predict these scores from phoneme sequences, achieving high accuracy despite the inherent limitations of a small dataset.

The dynamic shape generation function translated these scores into a continuum of geometric forms, ranging from smooth, curvy shapes to highly spiky, angular figures. The inclusion of a hybrid naming mechanism further linked the visual and auditory domains, providing a cohesive representation of sound symbolism. Preliminary user evaluations confirmed that the generated shapes align well with human intuitions, particularly at the extremes of the bouba–kiki spectrum.

Overall, this work not only contributes to the field of computational linguistics by offering a novel method for mapping phonemes to shapes but also opens up exciting possibilities for applications in design, education, and interactive systems. The integration of linguistic theory with deep learning has paved the way for a richer understanding of how abstract sensory experiences can be quantified and reproduced.

9. Future Work

While the results are promising, several avenues for future research remain:

9.1 Hybrid System Development

Integrate the deterministic rule-based methods with the deep learning approach to develop a fully hybrid system. Such a system could use rule-based algorithms as a preprocessing step, refining deep learning outputs for improved consistency and aesthetic appeal.

9.2 Dataset Expansion and Refinement

Expanding the dataset to include more real-world pseudowords and cross-cultural samples will likely enhance model generalizability. Crowdsourcing additional data through platforms such as Amazon Mechanical Turk could provide a richer set of human annotations, thereby improving both training and evaluation.

9.3 Advanced Model Architectures

Exploring more sophisticated architectures, such as Transformers or diffusion models, could capture even more subtle relationships between phoneme sequences and shape parameters. Additionally, generative adversarial networks (GANs) may be employed to produce even more diverse and realistic outputs.

9.4 Dynamic and Interactive Systems

Develop interactive, real-time applications that allow users to input words and immediately see the corresponding shape and hybrid name. Such systems could serve educational purposes or be applied in design contexts where intuitive sound–shape mappings enhance user experience.

9.5 Comprehensive User Studies

Future studies should involve larger and more diverse participant groups. Extended evaluations, including longitudinal studies, eye-tracking, or neuroimaging, could yield deeper insights into the cognitive processes underlying sound symbolism.

9.6 Ethical and Practical Considerations

As this work scales, ethical issues regarding data usage, model transparency, and fairness must be addressed. Guidelines for responsible deployment in commercial or educational settings should be developed.

10. References

Alper, M., & Averbuch-Elor, H. (2023). *Kiki or Bouba? Sound Symbolism in Vision-and-Language Models*. NeurIPS. PDF

Barker, H., & Bozic, M. (2024). *Forms, Mechanisms, and Roles of Iconicity in Spoken Language: A Review*. *Psychological Reports*, 131(1), 33–47.

Chen, Y. C., Huang, J., & Spence, C. (2019). *I know that Kiki is angular: Sound–shape correspondences in Mandarin speakers*. *Multisensory Research*, 32(1), 1–17.

Cho, P. (2005). *Takeluma: An Exploration of Sound, Meaning, and Writing*. UCLA MFA Thesis. PDF

Ćwiek, A., et al. (2024). *The alveolar trill is perceived as jagged/rough by speakers of different languages*. *Journal of the Acoustical Society of America*, 156(5), 3468–3481. Link

Fort, M., & Schwartz, J. L. (2022). *Resolving the Bouba–Kiki Effect Enigma*. *Scientific Reports*, 12, 19338. https://doi.org/10.1038/s41598-022-23623-w

Kim, S. H. (2020). *Bouba and Kiki Inside Objects: The Role of Visual Context*. *Perception*, 49(3), 253–267.

Matsuhira, C., et al. (2024). *Computational measurement of perceived pointiness from pronunciation*. *Multimedia Tools and Applications*, 83, 8213–8242. PDF

Maurer, D., Pathman, T., & Mondloch, C. J. (2006). *The shape of boubas: Sound–shape correspondences in toddlers*. *Developmental Science*, 9(3), 316–322.

Passi, P., & Arun, K. (2024). *The Bouba–Kiki effect is predicted by sound properties but not speech properties*. *Attention, Perception, & Psychophysics*, 86(1), 112–128.