

Bridging Dialectal Variation: A Phonetic Transcription Tool for Tamil

Ahrane Mahaganapathy¹, Sumirtha Karunakaran¹,
Kavitha Navakulan², Kengatharaiyer Sarveswaran¹

¹Department of Computer Science, University of Jaffna

²Department of Linguistics and English, University of Jaffna

Correspondence: ahrane@univ.jfn.ac.lk

Abstract

Phonetic transcription is vital for speech processing and linguistic documentation, particularly in languages like Tamil with complex phonology and dialectal variation. Challenges such as consonant gemination, retroflexion, vowel length, and one-to-many grapheme-phoneme mappings are compounded by limited data on Sri Lankan Tamil dialects. We present a dialect-aware, rule-based transcription tool for Tamil that supports Indian and Jaffna Tamil, with extensions underway for other dialects. Using a two-stage pipeline: Tamil script to Latin, then to IPA with context-sensitive rules, the tool handles dialect shifts. A real-time interface enables dialect selection. Evaluated on a 7,830-word corpus, it achieves 94.54% accuracy for Jaffna Tamil and is higher than other tools like eSpeak NG, advancing linguistic preservation and accessible speech technology for Tamil communities.

1 Introduction

Tamil is a classical South Dravidian language (Steever, 2019) spoken by nearly 80 million people across India, Sri Lanka, Malaysia, Singapore, and all over the world. Tamil has a proven literary history that spans more than two millennia, during which it passed through different phases of evolution: Old Tamil, Middle Tamil, and Modern Tamil, resulting in a dichotomy between literary Tamil, adhered to codified grammar (Lehmann, 2019) and spoken Tamil, which can vary significantly between regions and communities (Schiffman, 1998). This dialectal diversity, especially in low-resource languages such as Sri Lankan Tamil, presents significant challenges for linguistic documentation and computational modeling.

In the Tamil language, a single grapheme can correspond to multiple phonemic realizations, depending on its phonological context, morphological environment, or dialectal variation. This phe-

nomenon is known as one-to-many grapheme-to-phoneme (G2P) mapping.

Phonetic transcription, which provides a standardized representation of speech sounds, is fundamental in both linguistics and speech technology. The International Phonetic Alphabet (IPA) offers a cross-linguistic framework for transcribing phonemes (Association, 2015). However, Tamil phonology, marked by vowel length distinctions, consonant gemination, retroflexion, and context-sensitive plosive variation, demands transcription models that are context-sensitive and dialect-sensitive. These complexities are amplified in dialects like Sri Lankan Jaffna Tamil, which preserve conservative features no longer prominent in other Tamil dialects or varieties (Suseendrarajah, 1993).

There were extensive research exists on dialect-specific transcription in other languages such as BanglaIPA (Hasan et al., 2026) models six regional Bengali dialects using state-alignment algorithms to handle phonetic variability and out-of-vocabulary characters. TransDic (Garrido et al., 2018) supports multiple dialects of Spanish and Catalan by incorporating rules for primary and secondary phonetic phenomena, such as seseo and yeísmo. Similarly, Arabic (Hawasly et al., 2025), and American English (Károly, 2019) have been modeled with dialect-sensitive G2P systems.

In contrast, existing tools, such as eSpeak NG (Bernard and Titeux, 2021) *Anunaadam* (Virtualvinodh, 2014), offer IPA transcription for Tamil but lack support for dialectal nuances, a gap that disproportionately affects speakers of marginalized dialects. To address this gap, we present a rule-based phonetic transcription tool for Tamil, with particular emphasis on Sri Lankan Jaffna Tamil—a dialect that remains largely excluded from existing computational resources, despite being widely spoken not only in Sri Lanka but also by diaspora communities worldwide over the past

four decades.

Our key contributions include:

- **Dialect-Specific Modeling:** The first Tamil phonetic transcription system designed by considering dialectal variations, which captures key phonological features such as retroflexion, intervocalic alternation, and other allophones in vowels and consonants. The Language Technology Group (LTG) website ¹ contains a dedicated “Projects” section where the “Text to IPA transcription” tool and the relevant code can be found.
- **Linguistic Preservation:** By formalizing and digitizing features of a dialect, the tool supports phonetic documentation and speech technology development in low-resourced Tamil communities.
- **Scalability and Generalization:** The modular rule-based architecture of the tool enables adaptation to other low-resourced dialects of Tamil and similar phonetically rich languages.

Currently, most of the existing speech processing tools utilize ESpeak NG as the phonetic transcription tool, which does not support Tamil dialectal variations. By modeling Tamil’s phonological complexity and linguistic diversity, the proposed tool also contributes to inclusive NLP. It also supports applications in Text-to-Speech (TTS) and Automatic Speech Recognition (ASR).

2 Phonetics and Phonological Features of Tamil

Tamil consists of a rich phonetic system characterized by a diverse set of vowels and consonants that contribute to its unique sound structure. One of the important features of the language is its clear distinction between short and long vowels, as well as the phonemic role of consonant gemination (Kuno, 1958). The vowel inventory includes five basic vowels (/i/, /e/, /a/, /o/, /u/), each appearing in both short and long forms, resulting in ten contrastive vowel phonemes. This vowel length distinction plays a significant role in lexical differentiation (Kuno, 1958). Beyond these core phonemes, dialectal variation introduces additional phonetic complexity. For instance, Jaffna Tamil includes a

¹<https://www.ltg.jfn.ac.lk/>

broader set of vowel allophones, such as [ɛ], [i], and [ə], which are not typically found in standard varieties (Suseendirajah, 1993), as demonstrated in Table 1.

Consonants of the language include stops, nasals, fricatives, and liquids, with a notable distinction between voiced and voiceless obstruents. The language’s phonological rules are highly context-sensitive (Krishnamurthy, 1977), and consonant gemination further highlights the importance of phonological context (Balasubramanian, 1984; Heselwood, 2013). Krishnamurthy (1977) emphasizes the efficiency of Tamil’s orthographic system, wherein a single grapheme can represent multiple phonemic values depending on contextual cues. This characteristic reflects Zipf’s principle of least effort, which suggests that languages tend to minimize articulatory and cognitive load by reducing the number of distinct symbols while preserving expressive capacity (Cherry, 1971).

3 Tamil Phonetic Transcription

Tamil phonetics and phonology form a well-organized system that is essential for analyzing the linguistic structure. The phonological framework has several key features, including context-dependent phonetic variations, regional dialectal patterns, and the incorporation of lexical borrowings, all of which contribute to Tamil’s distinctive phonetic characteristics (Annamalai, 1979; Keane, 2004; Suseendirajah, 1993). This section provides a comprehensive examination of these aspects.

3.1 Context-Sensitive Pronunciations

Tamil phonology is characterized by significant context-sensitive variations, where the pronunciation of consonants changes based on their position within a word and the surrounding phonetic environment. A distinctive characteristic of Tamil phonology is the distribution of obstruents, particularly stops and fricatives. Voiced obstruents generally appear in word-internal positions, most often following nasal consonants. For instance, sequences such as a nasal followed by a voiced plosive, e.g., [ɲd̪] or [ɲj] are frequently observed within words. A few examples of context-sensitive pronunciation in Tamil, as handled by the proposed system, are presented in Table 2.

These variations present challenges in phonetic transcription tools, as they necessitate sophisti-

Phoneme	Allophone	Example
/a/	[ɛ]	அல்லி <i>alli</i> ‘water lily’ [ɛ l l i]
/a:/	[ɛ:]	நாய் <i>naai</i> ‘dog’ [nɛ:j]
/i/	[i]	நிறை <i>nirai</i> ‘weight’ [n i r a i]
/i:/	[i:]	நீளம் <i>neelam</i> ‘length’ [n i : l a m]
/u/	[u]	உப்பு <i>uppu</i> ‘salt’ [u p p u]
/e/	[ə]	எண் <i>en</i> ‘number’ [ə n]
/e:/	[ɛ:]	ஏணி <i>e:ni</i> ‘ladder’ [ɛ:n]

Table 1: Allophonic vowel realizations in Jaffna Tamil with example IPA transcriptions.

Grapheme	Allophone	Example
க்	[k]	கல் <i>kal</i> ‘stone’ [k ɛ l]
க்	[g]	கங்கை <i>gangai</i> ‘river Ganga’ [k a ŋ g a i]
க்	[x]	பகல் <i>pagal</i> ‘day time’ [p a x ɛ l]
ச்	[tʃ]	பச்சை <i>pachai</i> ‘green’ [p ɛ tʃ tʃ a i]
ச்	[dʒ]	மஞ்சள் <i>manjal</i> ‘yellow’ [m a n dʒ a l]
ச்	[s]	சட்டி <i>satti</i> ‘pot’ [s a tʃi]
த்	[t]	தம்பி <i>thambi</i> ‘younger brother’ [t a m b i]
த்	[d]	சிந்தனை <i>sindhanai</i> ‘thought’ [s i n̄ d ɛ n a i]
த்	[ð]	உதவி <i>uthavi</i> ‘help’ [u ð a v i]
ப்	[p]	பல் <i>pal</i> ‘tooth’ [p ɛ l]
ப்	[b]	பம்பரம் <i>pambaram</i> ‘spinning top’ [p a m b a r a m]
ப்	[β]	அபாயம் <i>abaayam</i> ‘danger’ [a β a: j a m]

Table 2: Some of the examples of one-to-many grapheme-to-phoneme mappings in Tamil across phonological contexts

cated modeling that can dynamically account for these positional and contextual changes. For instance, a single orthographic representation of a consonant might correspond to different phonetic realizations depending on its context, making accurate transcription complex without a context-aware system.

3.2 Dialectal Variations

Tamil exhibits considerable regional dialect variation across India, Sri Lanka, and other geographical regions. For instance, Jaffna Tamil is conservative and has unique vowel allophones that are generally absent in Indian Tamil dialects (Suseendirarajah, 1993). In contrast, urban varieties such as Chennai Tamil demonstrate significant phonetic and lexical influence from languages such as English, Hindi, Urdu, and Telugu (Smirnitskaya, 2018). These forms of variation underscore the need for transcription systems that are sensitive to dialectal variation across Tamil-speaking communities.

3.3 Loanword Influence

The integration of loanwords from Indo-Aryan languages such as words contain *Grantha* phonemes from Sanskrit and English has introduced new phonological elements into Tamil, altering traditional sound patterns. Even though, voiced plosives such as /b/ and /d/ are rare at the beginnings

of native Tamil words, they frequently occur in borrowed terms from languages like Sanskrit, Hindi, and English (Annamalai, 1979).

Therefore, the transcription systems must accommodate these non-native phonemes to accurately represent the evolving phonetic inventory. For instance, voiced fricatives such as /v/ and /z/, which are uncommon in the indigenous Tamil lexicon, are often present in borrowed vocabulary.

3.4 Notational Variances in Tamil Phonology

The notational system used to represent Tamil phonemes varies. Especially when using the International Phonetic Alphabet (IPA) to depict these sounds, ISO 15919 symbols are also used. The notational differences highlight the complexities of standardizing Tamil phonetic transcription, particularly when denoting dialectal variations. Some of the examples are listed in Table 3.

4 Methodology

According to Figure 1, the proposed Tamil Phonetic Transcription Tool adopts a rule-based architecture designed to preserve linguistic information while enabling dialect-sensitive phonetic realization. LLM based methods were not adopted because of the lack of annotated dialect specific corpora and their bias towards Indian Tamil varieties. The pipeline consists of (i) Tamil-to-

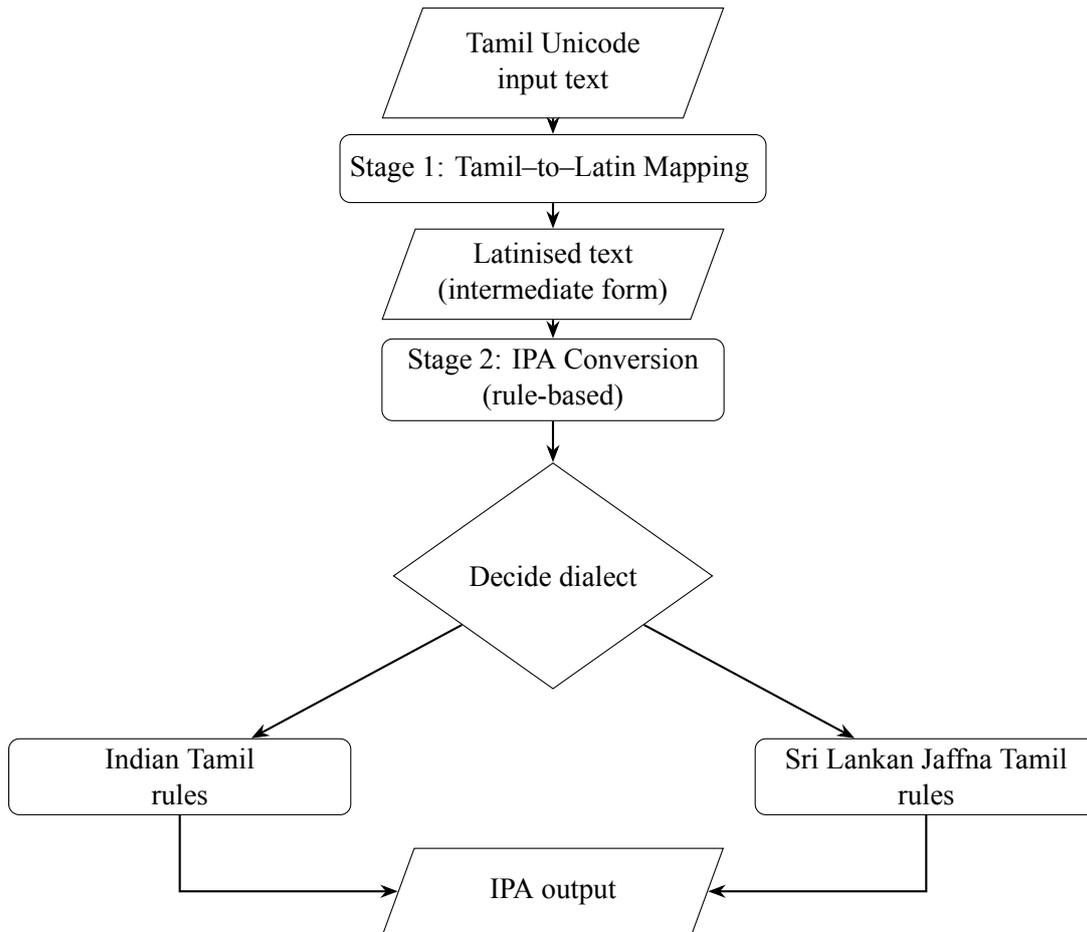


Figure 1: Pipeline for dialect-aware Tamil-to-IPA transcription.

IPA	ISO 15919
d	t
d̪	t̪
dz	c
j	y
l	ɭ
n	ɳ
ɲ	ɳ̠
ɳ	ɳ̠
t̪	t̪
tʒ	c
a:	ā
e:	ē
i:	ī
o:	ō
u:	ū

Table 3: Tamil Phonemes: IPA and ISO 15919 Equivalents

Latin grapheme mapping, followed by (ii) context-aware IPA conversion. It differs from eSpeakNG and *Anunaadam* as they do not cover dialectal vari-

ations in Tamil by adding dialect specific rule-set for Sri Lankan Jaffna Tamil.

4.1 Stage 1: Tamil-to-Latin Mapping (Information-Preserving Layer)

In the first stage, the `tam2lat()` module converts the Tamil script into an intermediate Latin-based representation. This mapping is strictly orthography-preserving: no segments are deleted, merged, or reinterpreted at this stage. The resulting representation remains structurally aligned with the original Tamil orthography, and all graphemic distinctions are explicitly retained. This conversion serves as a preprocessing step that facilitates the application of regular expressions, which are otherwise difficult to apply directly to Tamil Unicode characters. In addition, this intermediary representation enables easier extension of the system to other languages in the future.

4.2 Stage 2: IPA Conversion and Contextual Rule Application

In the second stage, the `ipa()` module maps the Latin intermediary to phonetic output through a sequence of context-sensitive transformations. These include intervocalic alternation of stops, nasal-stop voicing sequences, gemination stabilisation, and vowel length realisation. The transformations are applied only at this stage, ensuring that the phonemic contrasts encoded earlier are available for accurate contextual interpretation.

This stage introduces a dialect-selection mechanism that allows users to generate IPA transcriptions for different Tamil dialects. Currently, the system supports Indian Tamil (with *Anunaadam*-compatible phonetic behavior) and Sri Lankan Jaffna Tamil, with additional dialects being integrated in future extensions.

4.3 Dialect-Specific Rule Layer

The Jaffna Tamil mode extends the baseline rules by introducing a dedicated dialect layer implemented on top of the shared core rules. This layer contains 11 Sri Lankan Jaffna Tamil specific phonological and allophonic rules, including intervocalic situations and additional vowel allophones documented in dialectal descriptions (Suseendirajah, 1993). These rules are applied only when the Jaffna dialect is selected; the Indian Tamil option retains the *Anunaadam*-style realization. Other dialects can be extended in this layer by applying their dialect specific rules when their options are selected.

The Sri Lankan Jaffna Tamil specific rules introduced in the proposed system are listed in Table 4.

5 Results and Discussion

As the Indian Tamil rules are based on *Anunaadam* tool, the system was evaluated for the newly added Sri Lankan Jaffna Tamil dialect using a phonetic lexicon consisting of 7,830 words, including intervocalic situations, allophonic variations, *Grantha* phonemes, and other context-sensitive phonological contexts and the phonetic transcription accuracy was manually verified by a Sri Lankan native linguist. Out of them, 7403 were stated to be precise by the linguist, resulting in the word-level accuracy of 94.54%. A few cases in free variation in /t:r/ and /tt/, which are not captured in the proposed system, led to the reduction in ac-

curacy. The proposed system only results /t:r/ as output, does not cater to other variations in such instances. For example, in Sri Lankan Jaffna Tamil, the word முற்றிலும் *murrilum* ‘all inclusive’ exhibits phonetic free variation and is pronounced either as [mut:rilum] or [muttilum]. This variation, which occurs in fluent speech without altering semantic content, reflects a broader tendency to simplify consonant clusters, particularly the /tr/ sequence. These free variations are not captured on the proposed tool, resulting the reduction of accuracy.

A detailed analysis revealed notable phonetic differences between Sri Lankan Jaffna Tamil and Indian Tamil based on the provided dataset. One key distinction lies in the use of fricative versus plosive sounds: Indian Tamil employs the voiced velar fricative [ɣ], as in [aɣaŋga:ra], while Sri Lankan Jaffna Tamil substitutes it with the voiceless velar plosive [x], yielding [axaŋga:ra]. This substitution is a characteristic feature of Jaffna Tamil pronunciation. Vowel variation is another significant feature. Indian Tamil typically uses the low central vowel [a], as in [aɣandaiya:na], whereas Sri Lankan Jaffna Tamil often shifts to the mid-front vowel [ɛ], producing forms such as ax-endaiya:na.

For comparative purposes, eSpeak NG was also evaluated using the same corpus for the Sri Lankan Jaffna dialect. Its phonetic outputs were analyzed, enabling a direct comparison of transcription accuracy and dialectal adequacy between the proposed system and an established text-to-speech framework. Out of 7,830 words, 5,212 were judged to be correctly transcribed by the linguist. Hence, the word-level accuracy was 66.56%.

When eSpeak NG output is analysed, the grapheme க் is realized only as the context-sensitive variants [k] and [g]. In intervocalic positions, it is realized as [g], whereas in Sri Lankan Jaffna Tamil, the fricative [x] occurs in intervocalic positions (Suseendirajah, 1993).

In some instances, word-initial positions are also realized as g, for example, the word கேலி *kēli* ‘teasing’ is also transcribed by eSpeak NG as [gɛ:li] and குதி *kuti* ‘jump’ is transcribed as [gudi]. However, in Jaffna Tamil, [g], voiced dorso-velar plosive, occurs medially, in the sequence, and [k] voiceless dorso-velar plosive, occurs in word-initial position and in medial position following [r] (Suseendirajah, 1993).

eSpeak NG also transcribes ட as [b] in the

Target	Environment
/k/ to [x]	intervocalic, பகல் <i>pagal</i> ‘day time’ [p a x ε l]
/p/ to [β]	intervocalic, அபாயம் <i>abaayam</i> ‘danger’ [a β a: j a m]
/t/ to [ð]	intervocalic, உதவி <i>uthavi</i> ‘help’ [u ð a v i]
/k/ to [g]	only after nasal, கங்கை <i>gangai</i> ‘river Ganga’ [k a ŋ g a i]
/a/ to [ε]	before alveolar consonants /r/, /l/, /n/, அல்லி <i>alli</i> ‘water lily’ [ε l l i]
/a:/ to [ε:]	before /j/, /r:/, நாப <i>naai</i> ‘dog’ [nε:j]
/i/ to [i]	before retroflex consonants and /r/, நிறை <i>nirai</i> ‘weight’ [n i r a i]
/i:/ to [i:]	before retroflex consonants and /r/, நீளம் <i>neelam</i> ‘length’ [n i : a m]
/u/ to [u]	if not in first syllable, உப்பு <i>uppu</i> ‘salt’ [u p p u]
/e/ to [ə]	before retroflex consonants, /p/, /k/, /m/, /v/, /ŋ/ and /r/, எண் <i>en</i> ‘number’ [ə ŋ]
/e:/ to [ə:]	before retroflex consonants, /p/, /k/, /m/, /v/, /ŋ/ and /r/, ஏணி <i>e:ni</i> ‘ladder’ [ə:ŋi]

Table 4: Jaffna-Tamil specific rules with examples

word-final positions. Eg. குருவைப் *kuruvaip* ‘teacher.ACC’, [guruvaip] ends with [b]. In some instances, it results as [b] in the word’s first position. Eg. பத்திரமாக *pattiramaaka* ‘carefully’ is transcribed as [battirama:ga] through eSpeak NG. However, in Sri Lankan Jaffna Tamil, [b], voiced bilabial plosive occurs in the sequence of [mp], [np], and voicing of [b] is not so heavy as in Indian Tamil (Suseendirarajah, 1993).

நான் *naan* ‘I’ is transcribed as [na:n] which does not reflect a phonemic distinction between the orthographic forms ந and ன orthographies. In contrast, the proposed system distinguishes between the two (for example, [nān]). These distinctions are important for second-language learners to avoid confusion.

There are issues regarding vowel length from eSpeak NG, in some contexts words such as சுத்தியூட்ட *cahtiyuutta* [suttijutt] ‘to energize’ resulted in [ɥ] instead of a long vowel in the transcription. Whereas, the proposed system works well in this context.

In addition to that, eSpeak NG does not cover the phonological rules specific to the Sri Lankan Jaffna dialect listed in Table 4, including allophonic vowel variations. Even though the [k] to [g] transformation occurs after nasals, it does not adequately capture the Jaffna Tamil context, as in Jaffna Tamil, [g] occurs only after nasals and not in word-initial position.

The *Anunaadam* tool itself was not evaluated for comparative purposes, as our system extends *Anunaadam* to account for Sri Lankan Jaffna Tamil dialectal phonological phenomena, making a direct comparison with the tool irrelevant

6 Challenges in Tamil Phonetic Transcription

Automating phonetic transcription for Tamil involves navigating a range of linguistic complexities rooted in its phonological structure. Tamil exhibits strong context sensitivity in its sound system, where the realization of consonants and vowels depends heavily on their syntactic and phonetic environments (Krishnamurthy, 1977). For example, ப is [p] in பல் *pal* ‘teeth’ but becomes [b] in பம்பரம் *pamparam* ‘spinning top’ as mentioned in Table 2. These phenomena are addressed through syntax-directed transcription rules employing a one-symbol look-ahead mechanism, thereby providing a foundation for context-aware phonetic modeling.

Loanwords from Indo-Aryan languages, European and other languages came into contact due to regional contact and the migration of Tamil diaspora all over the world, adding another layer of complexity, often introducing phonemes not native to Tamil. As a result, transcription remains an ongoing process that must continually adapt to changing phonotactic patterns. For example [f] in ஃபோன் *fōn* ‘phone’. Foundational frameworks such as Caldwell’s Law (Caldwell, 1931) remain relevant, but modern transcription systems must adapt to a broader spectrum of linguistic inputs.

Altogether, these challenges underscore the need for transcription tools that are dialect-sensitive, context-aware, and capable of handling both standard and non-standard phonetic phenomena. Addressing them is critical for developing inclusive NLP systems.

7 Conclusion

This study presents a dialect-sensitive phonetic transcription tool for Tamil, called *ThamizhiIPA-*

Trans, addressing the phonological complexity and regional variability inherent in the language. Unlike existing tools, which often overlook dialectal nuances of low-resource languages such as Tamil, the proposed system is designed to model features specific to individual Tamil dialects, with a particular focus on Sri Lankan Tamil. A user-friendly interface allows real-time input and dialect selection, making the system useful for researchers, educators, and speech technology developers alike.

The developed system was evaluated on a lexicon of 7,830 words, and the outputs were assessed by a trained linguist, yielding an overall accuracy of 94.54%, which is substantially higher than the widely used eSpeak NG, which achieved 66.56% for the Sri Lankan Jaffna dialect. *ThamizhiIPA-Trans* can be accessed through the Language Technology Group (LTG) website². Future directions include extending the system to cover additional Tamil dialects in Sri Lanka and abroad and integrating prosodic features.

Limitations

Even though the proposed system demonstrates strong performance for Sri Lankan Tamil phonetic transcription, certain limitations remain. Some instances of free variation, speaker-specific pronunciation, code-mixed forms, and style-dependent phonetic shifts are not captured by the current rule set — addressing these may require a data-driven approach. The present approach models phonetic realization primarily at the segmental level, while prosodic phenomena such as stress, rhythm, and intonation remain outside the scope of this work. These limitations point to promising directions for future extensions, but do not detract from the central contribution of this study, namely, demonstrating that a dialect-aware, rule-based framework can effectively model key phonological processes in Tamil phonetic transcription.

Declaration on Generative AI

We used AI-assisted tools (e.g., ChatGPT) solely for language refinement and carefully rechecked and edited by the authors; all scientific contributions were carried out by the authors.

²<https://www.ltg.jfn.ac.lk/>

Acknowledgements

This research study was carried out under the DigSAL project, supported by the German Academic Exchange Service (DAAD) and funded by the Federal Ministry for Economic Cooperation and Development (BMZ) through SDG Partnerships (2024-2027).

References

- E. Annamalai. 1979. *Linguistic Complexity of Tamil Phonology*. Madras University Press.
- International Phonetic Association. 2015. *Handbook of the International Phonetic Association: A Guide to the Use of the International Phonetic Alphabet*. Cambridge University Press.
- T. Balasubramanian. 1984. Tamil Phonetic Characteristics: A Comparative Study. *Phonology Journal*.
- Mathieu Bernard and Hadrien Titeux. 2021. [Phonemizer: Text to Phones Transcription for Multiple Languages in Python](#). *Journal of Open Source Software*, 6(68):3958.
- Robert Caldwell. 1931. *Comparative Grammar of the Dravidian or South Indian Family of Languages*. University Press, London.
- C. Cherry. 1971. *On Human Communication*. MIT Press.
- Juan María Garrido, Marta Codina, and Kimber Fodge. 2018. TransDic, a public domain tool for the generation of phonetic dictionaries in standard and dialectal Spanish and Catalan. In *IberSPEECH*, pages 291–295.
- Jakir Hasan, Shrestha Datta, Md Saiful Islam, Shubhashis Roy Dipta, and Ameya Debnath. 2026. [BanglaIPA: Towards Robust Text-to-IPA Transcription with Contextual Rewriting in Bengali](#). *Preprint*, arXiv:2601.01778.
- Majd Hawasly, Hamdy Mubarak, Ahmed Abdelali, and Ahmed Ali. 2025. DialG2P: Dialectal Grapheme-to-Phoneme. Arabic as a Case Study. In *Proceedings of The Third Arabic Natural Language Processing Conference*, pages 466–471.
- Barry Heselwood. 2013. *Phonetic Transcription in Theory and Practice*. Edinburgh University Press.
- László Károly. 2019. An integrated dialect analysis tool using phonetics and acoustics. *Lingua*, 221:37–48.
- Elinor Keane. 2004. Illustrations of the IPA: Tamil. *Journal of the International Phonetic Association*, 34(1):111–116.
- E.V. Krishnamurthy. 1977. Automatic Phonetic Transcription of Tamil in Roman Script. *Proceedings of the Indian Academy of Sciences*, 86(6):503–512.

- Susumu Kuno. 1958. *Phonemic Structure of Colloquial Tamil*. *Journal of Linguistic Studies*.
- Thomas Lehmann. 2019. Old tamil. In *The Dravidian languages*, pages 81–103. Routledge.
- Harold F Schiffman. 1998. Standardization or re-standardization: The case for “standard” spoken Tamil. *Language in Society*, 27(3):359–385.
- Anna Aleksandrovna Smirnitskaya. 2018. Diglossia and Tamil varieties in Chennai. *Acta Linguistica Petropolitana. Труды института лингвистических исследований*, 14(3):318–334.
- Sanford B Steever. 2019. Introduction to the Dravidian languages. In *The Dravidian languages*, pages 1–44. Routledge.
- S. Suseendirajah. 1993. *Jaffna Tamil (1) Phonology and Morphology*. University of Jaffna Publication.
- Virtualvinodh. 2014. Anunaadam. GitHub repository. <https://github.com/virtualvinodh/anunaadam>.