Syllabification by Phone Categorization

Jacob Krantz, Maxwell Dulin, Paul De Palma

Gonzaga University
Spokane, WA

jkrantz@zagmail.gonzaga.edu

Mark VanDam

Washington State University
Spokane, WA

mark.vandam@wsu.edu

GONZAGA UNIVERSITY
School of Engineering & Applied Science

Abstract

Syllables play an important role in speech synthesis, speech recognition, and spoken document retrieval. A novel, low cost, and language agnostic approach to dividing words into their corresponding syllables is presented. A hybrid genetic algorithm constructs a categorization of phones optimized for syllabification. This categorization is used on top of a hidden Markov model sequence classifier to find syllable boundaries. The technique shows promising preliminary results when trained and tested on English words.

Background

Terminology

Phone: a unit of sound (*t* in the English **t**ip)

Syllable: a single segment of uninterrupted phones (*syl - la - bles*)

Syllabification: the process of breaking a word (a sequence of phones) into its corresponding syllables

Methods of Syllabification

1. **Rule-based**: Involves numerous handwritten rules about a given language. A prominent example would be the *tsylb* syllabification software based on Daniel Kahn's elaborate phonological algorithm [1].
2. **Probabilistic**: Statistical approaches based on training examples to provide learned insight. High order hidden Markov models (HMMs) and support vector machines (SVMs) have shown to perform this task at a state of the art level [2].

Training Advantage

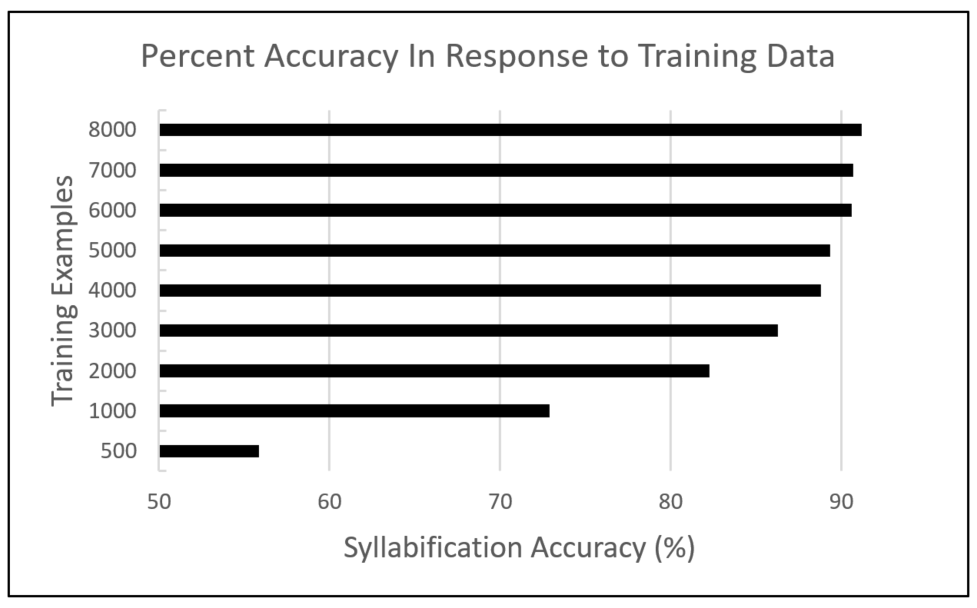

Figure 1

There are 54 phones in the IPA. With there being either a syllable boundary or not for each time step, the hidden state space is $54 * 2 = 108$. Using 12 phonetic categories, we reduce the hidden state space to $12 * 2 = 24$. Thus, the model achieves high accuracy with limited training data.

Conclusions and Future Work

Conclusions

1. **Our sequence classifier can accurately predict syllable boundaries at a word-level accuracy of 92.54% (using 10-fold cross-validated on CELEX).**
2. **Genetically-optimized phonetic mappings alongside the hidden Markov model show promise as a method of automatic syllabification.**

Future Work

- Test language independence against German, Dutch, and other languages.
- Investigate why certain phones pattern well in syllabification. Interestingly, the genetically-optimized categories do not pattern well with conventional, natural phonetic categorizations.
- Release the data and system of syllabification to benefit both researchers in linguistics and computational linguistics.

Method

Syllabification can be treated as a sequence classification problem. We use a version of the International Phonetic Alphabet called DISC to represent words as sequences of phones [3]. Before interacting with the model, these phones undergo a transformation based on a given table of one to many mappings. The phones on the left map to the category on the right:

$$[\ b\ ] \rightarrow A \qquad [\ \{, E\ ] \rightarrow B \qquad [\ s, n, t\ ] \rightarrow C$$
$$absent \rightarrow \{bsEnt \rightarrow BACBCC$$

These categories, enumerated as bigrams, form the input to the sequence classifier, a first order hidden Markov model (HMM)[4]. Given a bigram category sequence and a trained HMM, the Viterbi algorithm determines the most likely syllable boundary sequence. Syllables can then be trivially recovered.

Phonetic Categories: The Genetic Algorithm

An important consideration is how to create the table of phone-category mappings discussed above. We use a no-knowledge approach that is initialized with a random set of mappings. Adjustments are then made to find the ideal set of mappings such that the accuracy of syllabification using said mappings is maximized. We employ a genetic algorithm to optimally search the space of potential phone-category mappings. Our genetic algorithm includes the following components:

- **Sampling**: Stochastic Universal Sampling(SUS)[5]
- **Mating**: Scattered Crossover
- **Mutation**: Self-adaptive based on the standard deviation of the evaluation accuracy
- **Custom Step**: Takes the gene, or phone, involved in the most missyllabifications in the most fit member. The HMM is trained and tested with the phone permuted with every categorization to determine the best mapping for the individual phone.

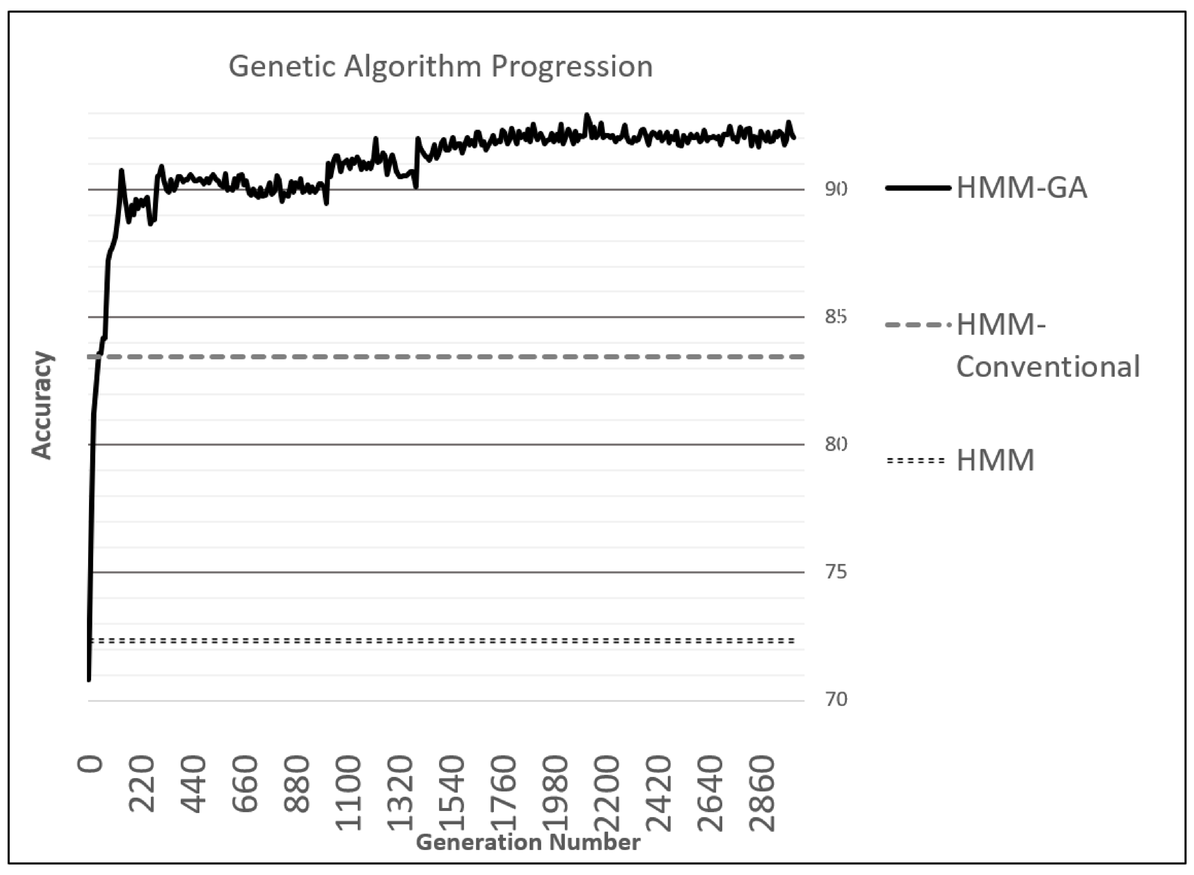

Figure 2: HMM-Conventional uses hand-crafted, natural phonetic categories. This is quickly surpassed by genetically optimized categories at just 83.45%.

References

[1] William Fisher. ''Tsylb syllabification software''. In: *National Institute of Standards and Technology: NIST* (1996).

[2] Susan Bartlett, Grzegorz Kondrak, and Colin Cherry. ''On the syllabification of phonemes''. In: *Proceedings of Human Language Technologies: The 2009 Annual Conference of the North American Chapter of the Association for Computational Linguistics*. Association for Computational Linguistics. 2009, pp. 308–316.

[3] R Harald Baayen, Richard Piepenbrock, and Leon Gulikers. ''Celex2''. In: *Linguistic Data Consortium, Philadelphia* (1996).

[4] Timothy Meekhof and Robert B Heckendorn. ''Using evolutionary optimization to improve markov-based classification with limited training data''. In: *Proceedings of the 7th annual conference on Genetic and Evolutionary Computation*. ACM. 2005, pp. 2211–2212.

[5] James E Baker. ''Reducing Bias and Inefficiency in the Selection Algorithm''. In: *Proceedings of the Second International Conference on Genetic Algorithms*. Vol. 206. 1987, pp. 14–21.