Progress Towards Eliciting Organized Phoneme Structures

Abstract

Phonological typology, a vital area within linguistic studies, examines the patterns and functions of sounds across the world's languages. This paper offers an overview of completed and ongoing experiments utilizing phonological representations, derived from typological databases, in speech processing tasks. It primarily focuses on two lines of inquiry motivated by the need to adapt speech technologies to low-resource languages and dialects. Initially, a framework is presented for evaluating the cross-linguistic consistency of phonological characteristics within multilingual phoneme inventories. Subsequently, an outline is given for a method that could potentially contribute to the development of future phoneme inventory induction systems, highlighting the crucial role of phonological typology in this process.

1 Introduction

The field of phonological typology investigates the distribution and functionality of sounds in languages globally. Typological databases are instrumental in making generalizations in this domain. These resources are valuable not only for creating probabilistic models of phonological typology but also for enhancing downstream multilingual NLP, speech technology, and language documentation efforts.

This paper summarizes our research involving phonological representations in speech processing, utilizing phonological typology databases. Despite the prevalence of end-to-end approaches in automatic speech recognition, text-to-speech, and speech-to-speech translation, the integration of precise phonological knowledge remains essential in various scenarios.

Our research is driven by the goal of extending speech technologies, which still rely on phonological representations, to under-resourced languages and dialects. We first review a framework designed to analyze the cross-linguistic consistency of phonological features in multilingual phoneme inventories obtained from cross-lingual typological databases. We then propose a preliminary method that may act as a foundational element in a future phoneme inventory induction system, emphasizing the significance of phonological typology in such an approach.

2 Multilingual Phoneme Inventories

Traditionally, a phoneme is defined as a theoretical concept specific to a single language. Applying phonemes and their feature encodings across languages presents a challenge: it's unclear whether all distinctive features (DFs) will be relevant or applicable in a multilingual phoneme inventory taken from a typological database. If DF representations were phonetic instead of phonemic, and acoustic rather than articulatory, one might anticipate a strong correlation between DFs and the acoustic signal. However, in practical multilingual contexts, these representations are frequently influenced by phonemic considerations due to the accessibility of phonemic inventories and transcriptions.

Our approach was straightforward: a phonemic contrast is deemed consistent across languages if it can be reliably predicted in a binary classification task on withheld languages. This problem involves a segment of a speech signal and a label (e.g., front vowel vs. back vowel). A classifier is trained on a

multilingual, multi-speaker dataset, excluding some languages for later assessment. In cases where cross-linguistic consistency was lacking, we enhanced the method by basing the representation on contextual phonological knowledge provided as DFs, excluding the contrast being tested.

Our experiments involved languages from the Dravidian, Indo-European, and Malayo-Polynesian families, with phoneme inventories sourced from a phonological database. The results are varied. An experiment designed to predict contrasts in unvoiced labial consonants between specific languages yielded reliable predictions across languages. Similar consistency was observed for contrasts between front and back vowels, as well as vowel height and continuant manner of articulation distinctions. Negative outcomes include the cross-lingual prediction of retroflex consonants between language families: a predictor trained on Dravidian languages cannot accurately predict retroflex consonants in another language, and vice versa. The detection of aspiration was similarly inconsistent. Incorporating other contrasts as contextual features did not result in significant improvement for these complex cases.

Our research is partly motivated by a persistent question: Can this methodology assess the cross-linguistic validity of existing phoneme inventories given available data? For example, among the Malayo-Polynesian languages studied, only one has retroflex consonants, acquired from loanwords from other language families. Different phoneme inventories exist for this language, one of which includes retroflex plosives, while another omits them. Determining which representation is superior in a multilingual pronunciation model remains an open issue.

3 Towards Phonology Induction

Our current research efforts are directed towards the induction of phoneme inventories for languages that lack the standard resources needed for speech model training. This task aligns with other work in zero-resource subword modeling. Existing unsupervised methods for discovering acoustic units derive acoustic-phonetic and latent auditory-like representations, but the typological accuracy of these representations is uncertain.

Our initial work with "universal" multilingual phoneme recognizers was not successful. This was mainly because the limited training data and the lack of language models to guide the search frequently led to unreliable phoneme inventory recovery even for closely related dialects, for instance, in trying to identify the phoneme inventory of one dialect after being exposed to two others. An improved strategy involves incorporating language identification and phonological typology into the phoneme recognizer. Using an accurate language identification model, the phoneme inventories of the most closely related languages can be employed to narrow down the potential phonemic hypotheses for a new, previously unseen language or dialect.

We are currently adapting the phonological contrast predictor methods from the previous section for phonology induction tasks. Several approaches for detecting phonemic features in continuous speech are known, some relying solely on signal processing and others that are model-based. At a basic level, the output of such predictors represents speech as parallel, asynchronous streams of articulatory features. More complex models that utilize the structure of articulation, feature geometry, and other correlations between features are also feasible. In these methods, cross-linguistic phonological databases are crucial for not only integrating various features into phonemes but also for validating which combinations of hypothesized phonemes are acceptable based on known phoneme inventories. Furthermore, additional phonological insights from other typological resources can be incorporated if they can be reliably extracted from the speech signal.