



LANGUAGE MODELING USING INDEPENDENT COMPONENT ANALYSIS FOR AUTOMATIC SPEECH RECOGNITION (MonPmOR6)

★ **Author(s) :** Raghunandan Kumaran (Clemson University, United States)
John Gowdy (Clemson University, United States)
Karthik Narayanan (Clemson University, United States)

★ **Abstract :** Conventional statistical language models such as N-grams are inadequate to model long distance dependencies in natural language. In this paper we propose a novel statistical language model to capture topic related long range dependencies. Humans have the inherent ability to identify long range dependencies in natural language. Given a set of related words humans can easily identify the context in which the set of words is occurring. It has been shown by many researchers that Independent Component Analysis (ICA) captures these kinds of dependencies better than any other formulation. Furthermore, ICA provides a topic decomposition that can be easily interpreted by humans compared to other models. This paper describes the development of a language model using ICA. The topic model is combined with a standard N-gram to produce the language model. The perplexity results obtained show that this language model is a viable language model for speech recognition purposes.