

I-SUEAT

A Special Problem
Presented to
the Faculty of the Division of Physical Sciences and Mathematics
College of Arts and Sciences
University of the Philippines Visayas
Miag-ao, Iloilo

In Partial Fulfillment
of the Requirements for the Degree of
Bachelor of Science in Computer Science by

CARPIO, Joecel Eman
GARCIA, Michael John
RABE, Jett Adriel

Francis Dimzon
Adviser

December 12, 2021

Chapter 1

Introduction

1.1 Insert Title here

With the rapid advancement of computer science and computational linguistics, numerous technologies are now being applied in various real-world situations. Among them is the interdisciplinary subfield known as speech recognition. Also known as automatic/computer speech recognition (ASR) or speech-to-text (STT), it utilizes computers in order to recognize and translate natural spoken language into text.

One of the key applications of automatic speech recognition is to transcribe speech documents such as talks, presentations, lectures, and broadcast news (Furui, et al., 2001, as cited in Furui, et al., 2004). A known challenge in speech transcription is that it can be quite taxing to retrieve and reuse speech documents if they are only recorded as audio. Although high recognition accuracy can be easily obtained for speech read from a text, such as anchor speakers' broadcast news utterances, technological ability for recognizing spontaneous speech is still limited (Furui, 2003, as cited in Furui, et al., 2004).

About 4500 languages exist in the world, but the majority of languages are spoken by less than 100,000 speakers; only about 150 languages (3%) have more than 1 Million speakers (Schultz, 2002).

Aklanon, which is often spelled as "Akeanon" by its local writers, is a dialect spoken by people located in the province of Aklan on the island of Panay in the Philippines. It somewhat varies with the dialects of neighboring provinces and islands and it belongs to a family of dialects whose ancestor might be proto-

West Visayan, which in turn was a member of the Malayo-Polynesian family of languages, to which such languages as Tagalog and Cebuano belong (De La Cruz & Zorc, 1968).

Aklanon/Akeanon is a specific language that is mostly exclusive to people who have lived in Aklan therefore most Filipino citizens wouldn't have familiarity with it. This would prove to be problematic since this would limit possible communications between locals and others. Additionally, there are barely any applications that are able to promote the language and get it out to the public. A speech-to-text recognition system would pave the way for the language to be recognized and appreciated.

This paper presents the development of a speech-to-text system that would be able to recognize Akeanon words with a decent accuracy rating. The system would be made using the open-source speech recognition toolkit "Kaldi". Kaldi is a speech recognition toolkit written in C++ and licensed under the Apache License v2.0. The system would be developed in a way such that the audio, acoustic and language data would be catered to the Akeanon dialect.

Chapter 2

Review of Related Literature

This chapter discusses... Hey guys, we need to put some filler text here to be honest. Right now its just random stuff but soon enough we should have something that we can place right here.

2.1 Automatic Speech technology

Automatic speech recognition (ASR) by machine has been a field of research for more than five decades. The industry has developed a broad range of commercial products where ASR as user interface has become ever more useful and pervasive [a1]. Automatic Speech Technologies are used daily in several applications and services. However, ASR technology has not reached a point where computers understand all speech in any acoustic environment or by any person. [a2]

The basic architecture of ASR involves four main components: signal processing and feature extraction, acoustic model (AM), language model (LM), and hypothesis search. Audio signals which are taken as input for signal processing and feature extraction components are enhanced and converted so it becomes suitable for acoustic models. The acoustic models are responsible for integrating knowledge about acoustics and phonetics. The language model processes the probability of a hypothesized word sequence from what it has learned from training. The hypothesis search component will combine the previous scores from both AM and LM and outputs a word sequence. [a3]

2.2 Speech Recognition Toolkits

Over the last few decades, commercial proprietary speech recognition systems have emerged such as AT&T Watson [b1], Microsoft Speech Server [b2], Google Speech API [b3] and Nuance Recognizer [b4]. These systems, however, are difficult to integrate into other software and they offered no control over the recognizer's features which led to the development of open-source automatic speech recognition systems.

In 2014, a comparative study was performed that involved training language and acoustic models for open-source ASR toolkits such as HDecode, Julius, pocketsphinx, Sphinx-4, and Kaldi. Their experiments showed an order of the evaluated toolkits regarding the ratio of effort to performance. The results showed that Kaldi outperformed all the other recognition toolkits when provided given out of the box training and decoding pipelines. The Sphinx also showed that it could provide training pipelines with a high possibility of achieving good results in a short amount of time. [b5]

2.3 Etymology of the Aklanon Language

While it is almost impossible to track the actual stages of development of the language, in a book published by De La Cruz and Zorc on Aklanon Grammar (De La Cruz & Zorc, 1968), the language comes from a long history of evolution of proto-languages which now turned out to be the present day Aklanon. The language itself has freely adopted English words. These words were simplified and have taken their own form of spellings. Aside from English words, there are also words borrowed from China, Spain and Japan that have been derived from to form Aklanon words. (Zorc, Prado, & Reyes, 1969)

2.4 Glottal Stop

The glottal stop presents a potential issue with speech to text recognition as native speakers of the Aklanon language, like most of its neighboring dialects, generally have a system wherein the words do not actually sound like they are spelled. These glottal stops have different rules concerning their positions such as initially before a vowel, medially between vowels, consonants or when a double glottal appears, and lastly the glottal at the final position (Zorc, Prado, & Reyes,

1969). Depending on how an individual might pronounce certain words, the results from speech to text recognition could possibly vary as different stress intonations on several words could result in a mismatch in recognition.