AUTOMATIC INDONESIAN WORD TEXT TO SOUND CONVERSION USING FINITE STATE TRANSDUCER OF DATA FLOW DIAGRAM IN SPEECH PROCESSING

Bernard Y. Tumbelaka, Haerudin
Faculty of Mathematics and Natural Sciences
University of Padjadjaran, Bandung
e-mail: bytumb@yahoo.com

Abstract

By combining machine learning techniques of finite state transducer (FST) in four-steps of data flow diagram level-1, we can develop conversion system of word text to sound automatically speech processing that recognizing and generating the pronunciation of words as correct as the pronounced words or inversely through a transformation-based learning. Our interest here is to design in device that is able to generate Indonesian word text to sound directly. We apply transducer composition to represent the combination of linguistic, phonetic and acoustic information at various levels as found. Generally, word text decomposition proceeds for each syllable recognizing must follow the rule of Indonesian word text decomposition by spelling words to syllables that consists of consonant and vowel sequences include in diphthong syllable identification. The syllables of phonetics are converted to acoustic composition same as the finite state transducer rules. A finite-state method based on leftmost longest-match replacement is presented here for segmenting words into syllables and for converting words into phonetics, which provides possibilities for developing transducer based on their conversion rules. Further, these word compositions are changed to sounds that are ready spoken. It is text to sound generated. It proposes is to know how the Indonesian syllables can be pronounced authentically as sounds orally. These sounds are recorded in wav-formats as audio-digital files. However, to store these wav-files we need an enormous hard-disk space, 3834 files with size of 244 MB.

Keywords: *speech processing, finite-state transducer (FST), data flow diagram level-1 (DFD Level-1), transformation-based learning, linguistics, phonetics, acoustics, wav-formats.*

1. INTRODUCTION

Automatic text into sound conversion is the most important application for speech recognition dealing with Indonesian word text [1,2], where the input may contain words. By combining machine learning techniques of finite state transducer (FST) in four-steps of data flow diagram level-1, we can develop conversion system of word text to sound through a transformation-based learning. This data flow diagram can be used to convert word text into sound automatically speech processing that recognizing and generating the pronunciation of words as correct as the pronounced words or inversely.

In particular, we use transducer compositions to represent the combination of linguistic, phonetic and acoustic information at various levels as found in a recognizer and a transducer [3]. Transducers can process from word texts of Indonesian language to sound applying the basic concept of finite state automata (FSA) as figured in figure 1, recognizing sentences by an acceptor M that specifies the language model, a transducer D from word sequences to context independent unit sequences, which is word decomposition in Indonesian language. Word decompositions is used to know how to pronounce Indonesian word texts, a transducer B that converts between sequences of context-independent and context-dependent phonetic units. A transducer A is mapping context-dependent phone sequences to acoustic observation sequences. This means that the words are assumed to be an oral language that needs to be explained briefly as often word decompositions in oral and in written is totally different. We may compose a generation task to acceptor 0 describing the acoustic observation sequence for the sound to be generated, that is, the likelihoods of different lexical transcriptions.

Usually, word text decomposition process for each syllable must follow the rule of Indonesian word decomposition by spelling words to syllables that consists of consonant (C) and vowel