

## INTISARI

# EVALUASI EMPIRIS DARI GATED UNIT RECURRENT NEURAL NETWORK SEBAGAI METODE UNTUK VOICE ACTIVITY DETECTION

Oleh

Muhammad Rifqi Fatchurrahman Putra Danar

12/335012/PA/15070

*Voice activity detection* (VAD), juga disebut sebagai *speech activity detection*, adalah proses pengidentifikasian secara otomatis, pada bagian kapan sinyal suara mengandung *speech* dan yang tidak mengandung *speech*. Semakin tinggi tingkatan *noise* dari suatu sinyal suara pengucapan, performa dari metode VAD-pun semakin buruk. Permasalahan tersebut menjadi objek penelitian dari VAD, dan salah satu metode yang memberikan performa yang baik ialah *Recurrent Neural Network* (RNN). Seperti yang sudah dilakukan pada penelitian sebelumnya, RNN dengan menggunakan *Long Short-Term Memory* (LSTM) mampu menunjukkan performa yang lebih baik dari metode *state-of-the-art* berbasis statistik. Performa dari LSTM-RNN juga pernah dievaluasi dengan model *deep learning* lain seperti *Convolutional Neural Network* (CNN) dan *Deep Neural Network* (DNN), dan menunjukkan performa yang lebih unggul dibandingkan dengan model *deep learning* tersebut. Selain LSTM-RNN, terdapat mekanisme *gating* lain dari LSTM-RNN, yaitu GRU-RNN. Penggunaan GRU-RNN belum pernah dibahas dan diimplementasikan pada VAD. Penelitian ini membahas tentang evaluasi performa secara empiris dari arsitektur RNN dengan mekanisme *gating* yaitu GRU-RNN dan LSTM-RNN sebagai metode untuk VAD.

GRU-RNN dievaluasi dan dibandingkan dengan LSTM-RNN dan tanh-RNN pada suara pengucapan bersih dan suara pengucapan dengan *noise* pada SNR 5 dB, 0 dB, dan -5 dB, menggunakan metrik *Area Under-ROC-Curve* (AUC) untuk deteksi *per-frame*. Hasil dari eksperimen menunjukkan bahwa GRU-RNN mampu mengungguli LSTM-RNN dan tanh-RNN pada suara pengucapan bersih, suara yang pekat oleh *noise* pada SNR 0 dB, dan -5 dB, dengan AUC 0.9684, 0.9142, dan 0.8504. Didapat bahwa GRU-RNN mampu mengungguli LSTM-RNN dan tanh-RNN pada suara pengucapan dengan *noise* tinggi, sehingga GRU-RNN dapat digunakan sebagai salah satu metode VAD.

Kata-kata kunci : *voice activity detection, recurrent neural network, gated recurrent unit, long short-term memory, deep learning.*

## ABSTRACT

### **AN EMPIRICAL EVALUATION OF GATED UNIT RECURRENT NEURAL NETWORK AS APPROACH FOR VOICE ACTIVITY DETECTION**

By

Muhammad Rifqi Fatchurrahman Putra Danar

12/335012/PA/15070

Voice activity detection (VAD) also referred to as speech activity detection, is an automatic identification process, on which part a speech signal contains speech and which does not contain speech. The higher the noise level of a speech sound signal, the performance of the VAD method is worse. That problem becomes the research object of VAD, and one of the methods that gives good performance is Recurrent Neural Network (RNN). As has been done in previous studies, RNN using Long Short-Term Memory (LSTM) is able to show a better performance than the state-of-the-art statistic based methods. The performance of LSTM-RNN has also been evaluated with other deep learning models such as the Convolutional Neural Network (CNN) and Deep Neural Network (DNN), and shows superior performance compared to the other deep learning model. In addition to LSTM-RNN, there is another gating mechanism from LSTM-RNN, called GRU-RNN. The use of GRU-RNN has not been discussed and implemented in VAD. This experiment discusses the empirical performance evaluation of the RNN architecture with gating mechanism of GRU-RNN and LSTM-RNN as a method for VAD.

GRU-RNN is evaluated and compared with LSTM-RNN and tanh-RNN on clean speech and noise robust speech with SNR 5 dB, 0 dB, and -5 dB, using Area Under-ROC-Curve (AUC) for detection on each frame. The result of the experiments show that GRU-RNN is able to outperform LSTM-RNN and tanh-RNN in clean speech, noise robust speech on SNR 0 dB, and 5 dB with AUC 0.9684, 0.9142, and 0.8504 respectively. It is found that GRU-RNN is able to outperform LSTM-RNN and tanh-RNN on noise robust speech, so GRU-RNN can be used as one of VAD method.

Keywords : voice activity detection, recurrent neural network, gated recurrent unit, long short-term memory, deep learning.