

2016 Pattern Recognition Association of South Africa and Robotics and Mechatronics International Conference, 30 November - 2 December 2016, Stellenbosch, South Africa

Speaker specific phrase break modeling with conditional random fields for text-to-speech

Johannes A Louw
Human Language Technologies Research Group
Meraka Institute, CSIR
Pretoria, South Africa
Email: jalouw@csir.co.za

Avashlin Moodley
Human Language Technologies Research Group
Meraka Institute, CSIR
Pretoria, South Africa
Email: amoodley1@csir.co.za

Abstract

In this paper we present a new cascading conditional random field based phrase break model for text-to-speech systems, trained on the speaker specific acoustic data that the text-to-speech voices are trained on. The training phase does not require any manually labeled phrase break tags, as these are derived directly from the speaker specific recordings used for building the synthetic voices. We present objective evaluations on various corpora, and show that the proposed model compares well with state-of-the-art data-driven phrase break models, with the added benefit of being in a unified framework.