1. Introduction

Audio Word2Vec

Jointly learn acoustic word segmentation and audio Word2Vec

Audio Word2Vec

Unsupervised training

2. Proposed Approach

Segmental Audio Word2Vec

Acoustic Word Segmentation Learning

Segmentation Precision/Recall and Segment Length

Example

Detected boundary

Auto-detected Word Boundaries

Spoken Term Detection Results

Application: Spoken Term Detection

3. Experiments

4 languages

Evaluation measure: mean average precision (MAP)

Oracle: using ground truth word boundary

DTW: frame-level dynamic time warping

Proposed approach significantly better than DTW

Obtained audio word representations did carry sequential phonetic information