

Accepted Manuscript

Effective Word Count Estimation for Long Duration Daily Naturalistic Audio Recordings

Ali Ziaei, Abhijeet Sangwan, John H.L. Hansen

PII: S0167-6393(15)30089-3
DOI: [10.1016/j.specom.2016.07.007](https://doi.org/10.1016/j.specom.2016.07.007)
Reference: SPECOM 2390

To appear in: *Speech Communication*



Please cite this article as: Ali Ziaei, Abhijeet Sangwan, John H.L. Hansen, Effective Word Count Estimation for Long Duration Daily Naturalistic Audio Recordings, *Speech Communication* (2016), doi: [10.1016/j.specom.2016.07.007](https://doi.org/10.1016/j.specom.2016.07.007)

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Effective Word Count Estimation for Long Duration Daily Naturalistic Audio Recordings

Ali Ziaei, Abhijeet Sangwan, John H.L. Hansen

*Center for Robust Speech Systems (CRSS),
University of Texas at Dallas (UTD), Richardson, Texas, U.S.A*

Abstract

The ability to count words in extended audio sequences allows researchers to explore characteristics of speakers (*i.e.*, leading, following, task responsibility, personal engagement), as well as the dynamics of two-way or multi-subject conversation scenarios. As such, counting the number of words spoken by a person, offers a rich information source for several applications such as health monitoring (*e.g.*, Autism, Parkinson's, Alzheimer's and etc), second language learning, or language development studies. However, developing robust word count systems that can achieve high performance with low computational cost is very challenging due to the uncertain and dynamic behavior experienced in audio recordings. In this study, we address the problem for large-scale naturalistic audio recordings based on a 100-day audio collection entitled, (*i.e.*, Prof-Life-Log). This corpus contains continuously recorded audio from one person using a mobile LENA audio recording device (LENA (2015)). The device captures audio for an entire workday which can last up to 16 hours. Our proposed framework to address word count consists of five main components, (i) **Speech Activity Detection(SAD)** to remove non-speech parts of the signal, (ii) **Speech Enhancement** to suppress the effects of background noise, (iii) **Primary vs. Secondary Speaker Detection** to remove secondary speaker segments, (iv) **Syllable Rate Estimation** to estimate the syllable rate for the primary speaker, and (v) **Linear Minimum Mean Square Error Estimation (LMMSE)** to find the linear mapping between syllable rate and word rate in spontaneous speech. In spite of the simplicity of the framework, it shows to be very effective in real scenarios with good performance on various datasets. As an indication of performance, the error of the framework for an entire 16-hours day audio file can be as low as 1% in terms of cumulative Word Count Error.

Key words: Word Count Estimation, Speaking Rate, Syllable Detection, Long Duration Audio, Prof-Life-Log

Download English Version:

<https://daneshyari.com/en/article/4977870>

Download Persian Version:

<https://daneshyari.com/article/4977870>

[Daneshyari.com](https://daneshyari.com)