

Accepted Manuscript

Towards Automatic Assessment of Spontaneous Spoken English

Y. Wang, M.J.F. Gales, K.M. Knill, K. Kyriakopoulos, A. Malinin,
R.C. van Dalen, M. Rashid

PII: S0167-6393(17)30454-5
DOI: <https://doi.org/10.1016/j.specom.2018.09.002>
Reference: SPECOM 2587

To appear in: *Speech Communication*

Received date: 8 December 2017
Revised date: 5 July 2018
Accepted date: 2 September 2018

Please cite this article as: Y. Wang, M.J.F. Gales, K.M. Knill, K. Kyriakopoulos, A. Malinin, R.C. van Dalen, M. Rashid, Towards Automatic Assessment of Spontaneous Spoken English, *Speech Communication* (2018), doi: <https://doi.org/10.1016/j.specom.2018.09.002>

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.



Towards Automatic Assessment of Spontaneous Spoken English

Y. Wang^{1,*}, M. J. F. Gales¹, K. M. Knill¹, K. Kyriakopoulos¹, A. Malinin¹, R. C. van Dalen², M. Rashid²

ALTA Institute / Department of Engineering, University of Cambridge, Cambridge, U.K.

Abstract

With increasing global demand for learning English as a second language, there has been considerable interest in methods of automatic assessment of spoken language proficiency for use in interactive electronic learning tools as well as for grading candidates for formal qualifications. This paper presents an automatic system to address the assessment of spontaneous spoken language. Prompts or questions requiring spontaneous speech responses elicit more natural speech which better reflects a learner's proficiency level than read speech. In addition to the challenges of highly variable non-native, learner, speech and noisy real-world recording conditions, this requires any automatic system to handle disfluent, non-grammatical, spontaneous speech with the underlying text unknown. To handle these, a strong deep learning based speech recognition system is applied in combination with a Gaussian Process (GP) grader. A range of features derived from the audio using the recognition hypothesis are investigated for their efficacy in the automatic grader. The proposed system is shown to predict grades at a similar level to the original examiner graders on real candidate entries. Interpolation with the examiner grades further boosts performance. The ability to reject poorly estimated grades is also important and measures are proposed to evaluate the performance of rejection schemes. The GP variance is used to decide which automatic grades should be rejected. Back-off to an expert grader for the least confident grades gives gains.

Keywords: Automatic assessment of Spoken English, Spontaneous speech, Pronunciation, Gaussian process, Rejection scheme

1. Introduction

There is a high demand around the world for the learning of English as a second language. Correspondingly, there is a need to assess the proficiency level of learners both during their studies and for formal qualifications. Given the vast number of non-native speakers combined with its overt status as the business language of choice nowadays, there are universally accepted tests such as International English Language Testing System (IELTS) and Test of English as a Foreign Language

(TOEFL). These tests often include listening, speaking, reading and writing sections that are marked by well-trained human examiners who assign a score based on a set of guidelines. To meet demand from English learners, the introduction of automatic graders for spoken language assessment would be beneficial especially for practice situations. The goal of an automatic grader is to assess language competence and provide scores reflecting the quality of the response from the candidates in a manner emulating the accuracy that could be achieved by a human grader. This could be fully automatic or combined with a human grader to boost the reliability of the system [1].

Compared to human graders, automated graders potentially perform more consistently and offer faster feedback times at a fraction of the marginal cost since the process of hiring and training new expert graders is costly and only offers a small increase in throughput. Figure 1 shows the architecture of a typical automatic assessment system for spoken language [2, 3, 4, 5, 6, 7]. Audio alone does not contain sufficient information to

*Corresponding author

¹Y. Wang, M. J. F. Gales, K. M. Knill, K. Kyriakopoulos, A. Malinin are with ALTA Institute / Department of Engineering, University of Cambridge, Cambridge, U.K., email: yw396, mjfg, kmk1001, kk492, am969@eng.cam.ac.uk.

²R. C. van Dalen and M. Rashid were with ALTA Institute / Department of Engineering, University Cambridge, Cambridge, U.K., during the course of the work.

³This paper reports on research supported by Cambridge Assessment English (CAE), University of Cambridge. Thanks to CAE for this support and providing access to the BULATS data.

Download English Version:

<https://daneshyari.com/en/article/10151554>

Download Persian Version:

<https://daneshyari.com/article/10151554>

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