Available online at www.sciencedirect.com

ScienceDirect

Speech Communication 75 (2015) 14–26

SPEECH COMMUNICATION

www.elsevier.com/locate/specom

Using automatic speech recognition to assess spoken responses to cognitive tests of semantic verbal fluency

Serguei V.S. Pakhomov [a,*], Susan E. Marino [a], Sarah Banks [b], Charles Bernick [b]

[a] Center for Clinical and Cognitive Neuropharmacology, University of Minnesota, Minneapolis, USA
[b] Lou Ruvo Center for Brain Health, Cleveland Clinic, USA

Received 22 July 2014; received in revised form 15 July 2015; accepted 22 September 2015
Available online 28 September 2015

Abstract

Cognitive tests of verbal fluency (VF) consist of verbalizing as many words as possible in one minute that either start with a specific letter of the alphabet or belong to a specific semantic category. These tests are widely used in neurological, psychiatric, mental health, and school settings and their validity for clinical applications has been extensively demonstrated. However, VF tests are currently administered and scored manually making them too cumbersome to use, particularly for longitudinal cognitive monitoring in large populations. The objective of the current study was to determine if automatic speech recognition (ASR) could be used for computerized administration and scoring of VF tests. We examined established techniques for constraining language modeling to a predefined vocabulary from a specific semantic category (e.g., animals). We also experimented with post-processing ASR output with confidence scoring, as well as with using speaker adaptation to improve automated VF scoring. Audio responses to a VF task were collected from 38 novice and experienced professional fighters (boxing and mixed martial arts) participating in a longitudinal study of effects of repetitive head trauma on brain function. Word error rate, correlation with manual word count and distance from manual word count were used to compare ASR-based approaches to scoring to each other and to the manually scored reference standard. Our study's results show that responses to the VF task contain a large number of extraneous utterances and noise that lead to relatively poor baseline ASR performance. However, we also found that speaker adaptation combined with confidence scoring significantly improves all three metrics and can enable use of ASR for reliable estimates of the traditional manual VF scores.

Keywords: Speech analysis; Cognitive testing; Automatic speech recognition; Verbal fluency; Speaker adaptation; Confidence scoring

1. Introduction

Tests of verbal fluency (VF) (Benton and Hamsher, 1994) are widely used in neurological, psychiatric, mental health, and school settings. There are two main types of VF tests – phonemic and semantic. The phonemic test (PVF) consists of naming as many words as possible in one minute that begin with a letter of the alphabet (e.g., A, F, or S). The semantic test (SVF) consists of naming words belonging to a semantic category (e.g., animals). These tests have been demonstrated to be useful for characterization of cognitive impairment due to a number of conditions including neurodegenerative disease (Henry et al., 2005, 2004; Henry and Crawford, 2004a), psychiatric diagnoses (Henry and Crawford, 2005), developmental disorders (Spek et al., 2009), drug toxicity or metabolic effects (Marino et al., 2012; Witt et al., 2013), as well as impairment due to traumatic brain injury or cardiovascular accidents (Henry and Crawford, 2004b, 2004c). In particular, contact sports such as boxing, mixed martial arts, football,

* Corresponding author at: 308 Harvard St. SE, Minneapolis, MN 55455, USA. Tel.: +1 612 624 1198.
E-mail address: pakh0002@umn.edu (S.V.S. Pakhomov).

http://dx.doi.org/10.1016/j.specom.2015.09.010
0167-6393/

and hockey are particularly well-known for high prevalence of repetitive head trauma which is a major risk factor for chronic traumatic encephalopathy (CTE), a devastating and untreatable condition that ultimately results in permanent disability and premature death (McKee et al., 2013). Athletes with prior exposure to head trauma show significant declines in verbal fluency performance among other types of cognitive impairment (Tremblay et al., 2013).

While clinically useful, VF tests are currently administered manually and are too cumbersome for wide adoption on a large scale in fast-paced and overburdened healthcare systems. Furthermore, manual VF testing is also prone to scoring subjectivity and variability, and cannot be easily self-administered, which limits its applications in large and/or longitudinal investigations of cognitive biomarkers of neurodegenerative disease. One of the goals of developing automated VF testing is to enable easy and non-threatening long-term monitoring of cognitive performance in an attempt to detect early subtle cognitive changes that may warrant a more in-depth clinical assessment. Following the desiderata for computerized VF testing (Kemper and McDowd, 2008), we propose to address existing limitations of VF tests by automating their administration and scoring. Our approach consists of using automatic speech recognition (ASR) technology applied to the speech collected during VF testing to estimate an approximate count of "legitimate" words produced during the VF task.

A typical administration of VF testing includes instructing the subject to restrict his or her responses to the specified stimulus (letter or category) and avoid using proper names. A subject's test-taking behavior is commonly evaluated both by the total number of correct responses produced during the task, and by an analysis of the subject's errors or incorrect responses. Under ideal circumstances, given all correct responses, an individual's speech produced on these tasks would consist of 1-, 2-, or 3-word phrases denoting the concepts relevant to the test (e.g., animals) separated by silent pauses without any repetitions, disfluencies, comments, or other extraneous utterances and noise. However, in reality, response errors and noise are common and subjects' speech often contains events that should not be included in the calculation of the total test score, which presents a significant challenge for automation (Miller et al., 2013).

Computerized administration and scoring of VF tests is a promising area in which computational linguistic and computerized speech processing approaches can make a significant contribution. While ample evidence exists to show clinical usefulness of manually administered VF tests, in order to extend this body of evidence to automated approaches, it is necessary to demonstrate that automated approaches provide acceptable estimates of the manual VF assessments in a variety of populations and environments (Bauer et al., 2012). Miller and colleagues successfully experimented with using an ASR-driven interactive voice response system to administer and score VF tests (Miller et al., 2013). However, using ASR for automatic scoring of VF tests remains a largely unexplored area and needs to be investigated further. To our knowledge, the use of various ASR techniques such as acoustic speaker adaptation and confidence scoring of ASR output have not been investigated in relation to optimizing automatic VF scoring.

Previous work shows that unsupervised speaker adaptation of acoustic models used in ASR can significantly improve recognition accuracy even with very limited (as little as 11 s) amounts of available adaptation data (Leggetter and Woodland, 1995; Wang et al., 2007) and is useful in a number of specific applications including dialog act segmentation (Kolář et al., 2010) and language tutoring systems (Ohkawa et al., 2009). Similarly to speaker adaptation, the use of confidence scoring in general-purpose ASR applications has also been well documented (see Jiang (2005) for a review), although the findings with respect to confidence scoring are mixed. Confidence scoring approaches tend to be highly application specific (Zeljkovic, 1996) and lack a systematic way of determining the most optimal confidence threshold (Bouwman et al., 1999). In the current study, we apply confidence scoring to post-process ASR output from VF responses in the context of using a language model with a highly constrained vocabulary in an attempt to leverage the fact that any speech that does not contain an item from the semantic category of interest would get a lower confidence score and thus may be reliably filtered out. The objective of the current study was to experiment with speaker adaptation and confidence scoring as applied to the specific task of automated VF assessment and to validate these approaches on an 'animal' verbal fluency test in a sample of cognitively normal individuals.

2. Materials and methods

2.1. Animal name recognition system

Standard ASR approaches (described in the next section in more detail) were used to create a system designed specifically to yield a score that represents how many animal names the speaker was able to produce in response to the VF task. The resulting animal name recognition system consisted of the ASR decoder with a specially trained animal fluency language model and a speaker adapted acoustic model, as well as a set of post-processing filters. The post-processing filters were designed to account for the spurious words that may appear in the raw ASR output. Some of these spurious words represent errors produced by the ASR decoder; however, many of these words represent animal naming errors – non-animal names that were recognized correctly by the ASR decoder. The latter type of errors should not be counted against the accuracy of the ASR system but should be counted against the accuracy of the animal name recognition system. Thus the SVF score estimate produced by the animal naming recognition system represents the number of words that