

Automatic administration of semantic verbal fluency tests for Portuguese

João Costa, Ana Pereira, Liliana Ferreira

Fraunhofer Portugal AICOS

<https://doi.org/10.36505/ExLing-2019/10/0015/000377>

Abstract

Verbal fluency tests are quick and flexible tests used in the area of Neuropsychology to evaluate executive and language functions of subjects. An automatic modular pipeline for the analysis of verbal fluency tests for the semantic category of animals is proposed in this document. The system was developed for European Portuguese using test audio recordings as input. Google Cloud Speech-to-Text is used to perform automatic speech recognition. *fastText* word embeddings and phonemic transcriptions are applied to automatically evaluate temporal, phonemic, and semantic clusters produced. The proposed architecture was evaluated using 164 animal category tests, performed by Portuguese elderly subjects. Correlations were found between manual and automatic extracted features, such as the number of correct words produced ($\rho=0.58$). Feature extraction efficacy was found to be dependent on background noise levels during recording.

Key words: verbal fluency tests, natural language processing, automatic speech recognition

Introduction

Verbal Fluency Tests (VFT) are frequently used to assess verbal and executive functions of subjects, as a standalone cognitive test or as part of a neuropsychological test battery. These tests consist in the production of as many words as possible, with respect to a given rule. For semantic VFTs, subjects are asked to produce words that belong to the same semantic category within a one minute time window; animal names are frequently used for this task.

In this work, a fully automatic and modular architecture for VFT administration is proposed. The developed pipeline was constructed for use with semantic VFT (animal names) in European Portuguese. With the proposed system, recorded audio from a VFT can be automatically transcribed, filtered, and processed, thus quickening and improving analysis of these tests, with the extraction of clinically relevant features.

Automatic administration of verbal fluency tests

The pipeline of the proposed system for automatic administration and analysis of VF tests is shown in Figure 1. The system was implemented in Python 3.6,

and uses as input audio recordings of VFT execution. The developed solution can extract 20 metrics from a single test.

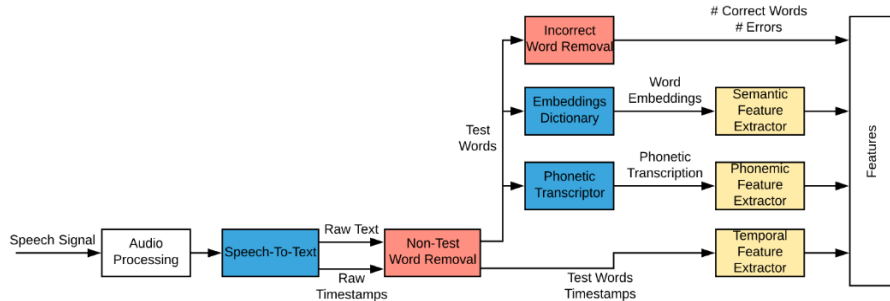


Figure 1. Overview of developed pipeline for automatic transcription and analysis of VFTs.

Speech-to-text

To perform Automatic Speech Recognition (ASR), the Google Cloud Speech-to-Text (Google Cloud 2019) service is used. This module converts each audio recording to a list of transcribed words, in European Portuguese, with the corresponding word timestamps in the audio signal.

Feature Extraction

Using the test transcript obtained after word filtering, a total of 20 features are extracted for each VF test, namely, the number of correct words produced, the number of errors, six temporal features, six semantic features, and six phonemic features.

Temporal features are derived from the predicted temporal clusters, defined using the DBSCAN clustering algorithm (Xiaowei 2002), from the word timestamps provided by Google Cloud Speech-to-Text.

Phonemic features are derived from the predicted phonemic clusters, which are based on the phonemic similarity between VFT exemplars. The phonemic similarity of a word pair is calculated using the Levenshtein distance between their phonetic transcriptions, provided by Infopédia (Porto Editora 2019).

Semantic features are derived from the predicted semantic clusters, which are based on the semantic similarity between VFT exemplars. The semantic similarity of a word pair is calculated using the cosine distance between their corresponding *fastText* deep word embeddings (Grave 2018).

Data collection

To validate the developed solution, 164 semantic VFT (animal names) recordings were collected with a *smartphone* from Portuguese elderly subjects. Recordings were performed in the scope of the clinical trial performed by Amorim & Susano (2017). Manual transcription was performed by the first

author, annotating all test exemplars (including errors, repetitions, and word variations).

This study was approved by the competent bodies *Departamento de Ensino Formação e Investigação*, *Comissão de Ética*, and *Conselho de Administração* of Centro Hospitalar do Porto, approval number 253/16 (216-DEFI/ 205-CES).

Results

ASR Performance

The Verbal Fluency Word Error Rate (VFWER) (König 2018) is used to evaluate the performance of the selected Speech-to-Text system (Figure 2). Performance of automatic transcription of VFTs was found to be highly dependent on background noise levels present in recordings, due to competing sounds and voices that obfuscate the speech of the subject performing the VFT.

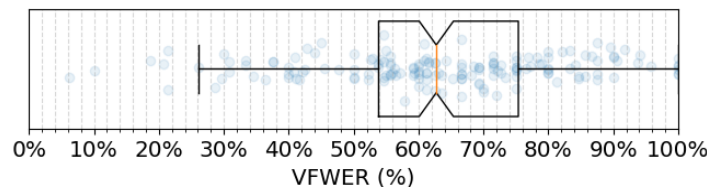


Figure 2. Verbal Fluency Word Error Rate distribution.

Feature Extraction

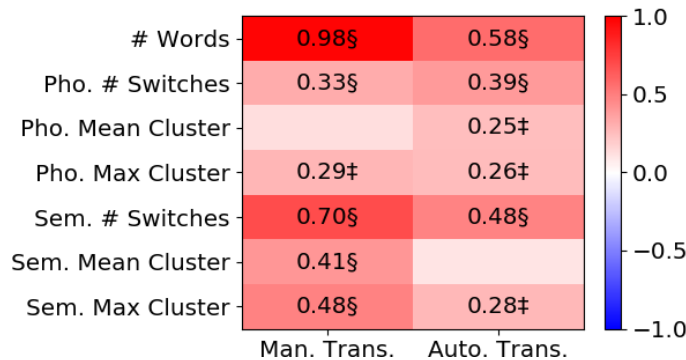


Figure 3. ρ of analogous feature pairs for manual and automatic transcriptions of VFTs ($\ddagger p < 0.001$; $\S p < 0.0001$).

Spearman's ρ was used to determine correlation between manually and automatically determined analogous feature pairs, using both manual and automatic transcriptions of VFTs (Figure 3). The high levels of VFWER lead to

weaker correlations for automatic transcripts of VFTs. Different approaches used for manual and automatic clustering may contribute for lower correlations, with the former based on strict rule sets, and the latter on more generalised automatic approaches by word embeddings and phonetic transcript similarities.

Conclusions

The new method of VFT analysis proposed in this document can potentially be used in the future in clinical settings by health professionals. Since extracted metrics have predictive power and can be used as screening tests for neurocognitive disorders, automatic VFTs can become a very powerful, easy to implement, and quick to apply tool to aid screening and diagnosis. Additionally, the modular design used for the implementation of the tool facilitates its adaptation to other tasks and languages.

Acknowledgements

The authors would like to thank Dr. Pedro Amorim, Dr. Maria João Susano, Dr. Francisco Seixas and Mariana Dias of Centro Hospitalar do Porto, and Prof. Anibal Ferreira of Faculdade de Engenharia da Universidade do Porto for their interest, availability, and data sharing.

References

- Amorim, P., Susano, M. J. 2017. Clinical Trial: Pre-operative Neurocognitive Disorder and Low Near-infrared Spectrometry is Associated with Postoperative Delirium. <https://clinicaltrials.gov/ct2/show/NCT03171766?id=NCT03171766>
- Google Cloud. 2019. Cloud Speech-to-Text - Speech Recognition | Cloud Speech-to-Text API. <https://cloud.google.com/speech-to-text/>
- Grave, E. et al. 2018. Learning Word Vectors for 157 Languages.
- König, A. et al. 2018. Fully automatic speech-based analysis of the semantic verbal fluency task. *Dementia and Geriatric Cognitive Disorders*, 45(3-4):198–209.
- Porto Editora. 2019. Infopédia - Dicionários Porto Editora. <https://www.infopedia.pt/>
- Troyer, A.K. et al. 1997. Clustering and switching as two components of verbal fluency: Evidence from younger and older healthy adults. *Neuropsychology*, 11(1), 138–146.
- Xiaowei Xu et al. 2002. A distribution-based clustering algorithm for mining in large spatial databases. *Proceedings 14th International Conference on Data Engineering*, 324–331.