

AN EMPIRICAL ANALYSIS OF STUTTERING AND STAMMERING SPEECH (SS-S) USING ERGONOMICS

A.Lalitha Venkatesan¹

¹Associate Professor, Department of CSE, Geethanjali College of Engineering, Hyderabad

Abstract

This paper describes about the state of art of speech recognition which will be helpful in identifying the stuttering and stammering speech of a person (disability). Every speech that is spoken should have proper quality tone and utterance but stammering speech is exceptional, as it lacks clarity. Successful communication depends on intonation, stress, rhythm where those attributes are very less in SS speech. To overcome this issue, there are few speech therapy process which are carried out (medically) those are less responsive and long term training is needed and cost effective too. To deal with this problem an ergonomic approach used in form of Mental Speech Model (MSM) to enhance the speech quality of every utterance of word. To create a MSM, a deep learning technique is needed which involves with Deep Neural Networks (DNN) over the SS speech. This model will have two repositories, one speech of normal person and other speech of abnormal person (SS-S) which is compared and matched to reflect the quality of speech utterance. The ergonomic approach that is used here can also act as a Psycho-Linguistic Approach (PLA) to train the system to improvise the speech quality to overcome the difficulties with stuttering and stammering speech. The ergonomic approach or PLA is effective in terms of speech process and cost.

Keywords— stuttering, stammering, speech quality, DNN, MSM, PLA

1. INTRODUCTION

Speech Recognition for direct communication between humans and computers, through HCI (Human Computer Interaction). Speech is the most natural communicative medium for humans to share their thoughts and views. Generally giving commands to a computer or other device, especially in environments where keyboard input is awkward or impossible where speech comes into picture which can be very interactive way of giving input. There are two types of system in SR they are "Speaker Independent"(no training) and "Speaker Dependent"(training). Speech recognition can play an important role in deep learning a cognitive learning. There is world-wide industry adoption of a variety of deep learning methods in designing and deploying speech recognition systems which are costly. DNN based on context dependent HMM states constructed by decision trees were adopted using human engineering(ergonomics) which can make SS-S to be more recognizable and understandable. Stuttering also known as stammering is a speech disorder in which the flow of speech is disrupted by involuntary repetitions and prolongations of sounds, syllables, words or phrases, and involuntary silent pauses or blocks in which the stutterer is unable to produce sounds. Causes for anxiety, low self-esteem, nervousness, too many workloads, unnecessary thinking, depression, bad voice quality and stress. Symptoms resembling stuttering are Asperser's syndrome, cluttering, Parkinson's speech, essential tremor, palilalia, spasmodic dysphonia, selective mutism and social anxiety, mental retardations (medically proven). Stuttering is less prevalent in deaf and hard-of-hearing individuals, and stuttering may be improved when auditory feedback is

altered, such as masking, delayed auditory feedback (DAF), or frequency altered feedback.

The basic understanding of cognitive, physical, behavioral, physiological, social, developmental, and motivational aspects of human performance can yield more to SS-Speech to enhance training, selection, communication to improve human-system interfaces or HCI. In further sections 2, 3, 4 we can better understand the process of empirical analysis on stuttering and stammering speech using ergonomics.

2. EXISTING WORK

In existing work there are various techniques like band equalization and hidden Markov models are used and to validate the meaningfulness of the spectrum of the filtered TSSP along with the usefulness of the alternative filter structure and the Slepian filters. Here design method for speaker-independent word recognition tasks are carried out. There also exists a Novel approach to perform simple statistical User Simulation (US) for this task, which measures the likelihood that the user would say each hypothesis in the current context. The method achieved a significant relative reduction of Word Error Rate (WER) and identifying the likelihood for each hypothesis where these methods seems to be very time consuming and sometime result in fuzzy values which makes it very hard to recognize the speech concerned. The word error rate can be computed as:

$$WER = S+D+I/N$$

Where

- S is the number of substitutions,

- D is the number of deletions,
- I is the number of insertions,
- C is the number of the corrects,
- N is the number of words in the reference ($N=S+D+C$)

Word error rate (WER) is a common metric of the performance of a speech recognition or machine translation system which also depends on certain work elements of speech recognition are sound cards, microphones, computers/processors and speech spectrograph.

3. PROPOSED WORK

The key functionality of this research is about enhancing the speech of stammering voices using DNN(Combination of Deep learning and Neural Networks).One fundamental principle of deep learning is to do with hand-crafted feature engineering which can also be called as human engineering to create accurate result on the research performed(use raw features). This principle was first explored successfully in the architecture of deep auto encoder on the "raw" spectrogram or linear filter-bank features, showing its superiority over the Mel-Cepstral features which contain a few stages of fixed transformation from spectrograms. The true "raw" features of speech, waveforms, have more recently been shown to produce excellent larger-scale speech recognition results reducing the noise or .

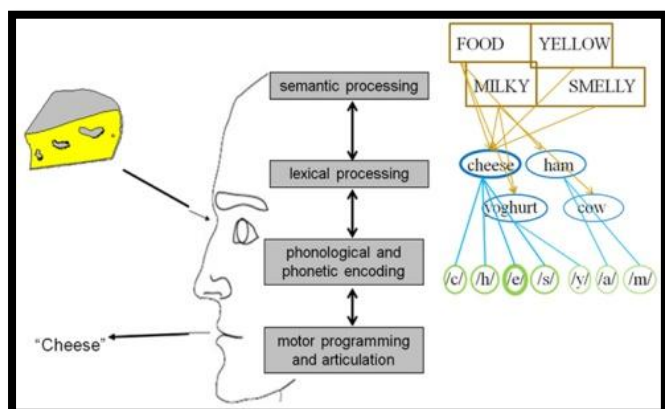


Fig 1: Look, Feel, Process, Store, Spell (LFPSS)

The above LFPSS method is used to process the given speech in ergonomic fashion by giving highest priority to the SS-Speech. Studies utilizing positron-emission tomography (PET) have found during tasks that invoke disfluent speech, stutterers show hypo activity in cortical areas associated with language processing, such as Broca's area, but hyperactivity in areas associated with motor function. There are several approaches for speech recognition like template based approach, Rule based approach, and statistics based approach. Speech Recognition through deep learning gives ergonomic benefits. Use of some euphemism like 'stress analyzer' or 'emotion analyzer', Diogenes Digital Voice Stress Analysis application was originally used in determining attempts. But here PLA approach is used to train the SS speech using phonetic encoding. This can be applied at various levels to

analysis the given stuttering and stammering voice. It is also worth pointing out that while physiological tremor has mostly been found and measured in muscles under static tension, the speech organs responsible for voice production are typically in constant motion. There are various algorithms (Tandem algorithm, MNN) that are used where creating a Mental Speech Model (MSM) along with Psycho-Linguistic Approach (PLA) can help in generating some phoneme specific models over human engineering or ergonomics.

4. EXPERIMENTATION

An empirical analysis was performed, to evaluate accuracy and speed of stuttering and stammering speech where accuracy is rated with word error rate (WER), speed with the real time factor. The inference seems to be speaker dependence or independence, discontinuous or continuous speech but in general identifying the correct phoneme from the person who as the disability of speech or vocally challenged still difficult but proper deep learning of the speech can make the system efficient when proper spectral matching of voice is done and breakage of the spoken word into proper phoneme means a lot. And this can be achieved only by LFPSS which is same as MSM along with PLA.

Steps Involved

1. Collect the samples (speech of normal person and speech of stammering) by audio recording and utterance detection.
2. Store them in different repositories to fetch when comparison is needed?
3. Analyze their ergonomic value using PLA
4. Pre-filtering by normalization.
5. Framing and windowing into usable format.
6. Apply the rules needed to design DNN.
7. Comparison and matching the utterance using mental speech Model (MSM)
8. Follow the PLA (Psycho-linguistic approach) to train MSM.
9. Enhanced speech quality is attained using the MSM that is created thus reducing the stuttering stammering speech.

5. CONCLUSION&FUTURE ENHANCEMENTS

It is just an approach suggested to enhance stuttering and stammering speech mainly to increase the speech quality, thereby reducing cost that is spent on various speech therapies. Suggestions are made to streamline the speech of normal person and speech of vocally challenged person using PLA along with DNN concepts. Future work depends on the implementing the concept of PLA with fewer stuttering and stammering speech samples, to justify the speed and accuracy.

REFERENCES

- [1] "Fundamentals of Speech Recognition". L. Rabiner & B. Juang. 1993. ISBN: 0130151572.
- [2] Memory System Design Space Exploration for Low-Power, Real-time Speech Recognition, Rajeev Krishna, Scott Mahlke, and Todd Austin.Proc. 2004

- Intl. Conference on Hardware/Software Codesign and System Synthesis (CODES-ISSS) Sep. 2004, pp. 140-145.
- [3] Insights into the Memory Demands of Speech Recognition Algorithms, Rajeev Krishna, Scott Mahlke, and Todd Austin. Proc. ACM/IEEE 2nd Workshop on Memory Performance Issues (WMPI) May 2002.
- [4] Howell, P. (2011). Supplementary material to Nature correspondence (February 24th, 2011) Howell, P., Anderson, A., Bartrip, J. & Bailey, E. (In Press). Comparison of acoustic and kinematic approaches to measuring utterance-level speech variability. Journal of Speech, Language and Hearing Research.
- [5] Howell, P., Davis, S. & Bartrip, J. (In Press). The UCLASS archive of stuttered speech. Journal of Speech, Language and Hearing Research.
- [6] Anderson, A., Lowit, A. & Howell, P. (2008). Temporal and spacial variability in speakers with Parkinson's disease and Friedreich's ataxia. Journal of Medical and Speech Language Pathology, 16, 173-180.
- [7] Geoffrey Hinton, Li Deng, Dong Yu, George Dahl, Abdel-rahman Mohamed, Navdeep Jaitly, Andrew Senior, Vincent Vanhoucke, Patrick Nguyen, Tara Sainath, and Brian Kingsbury, Deep Neural Networks for Acoustic Modeling in Speech Recognition, April 27, 2012
- [8] Gregor McGlaun *, Frank Althoff *, Hans-Wilhelm Rühl +, Michael Alger °, Manfred Lang *, A generic operation concept for an ergonomic speech MMI under fixed constraints in the automotive environment..

BIOGRAPHY



A.LALITHA VENKATESAN is presently working as an Associate Professor in the department of CSE at **Geethanjali college of Engineering and Technology**, Hyderabad with the work experience of 8+ years and certified software tester too.