Automatic Speech Intelligibility Assessment in Dysarthric Subjects

Ayush Tripathi, Swapnil Bhosale, Sunil Kumar Kopparapu
TCS Research and Innovation - Mumbai, India
Email: {t.ayush, bhosale.swapnil2, sunilkumar.kopparapu}@tcs.com

***Abstract*—Dysarthria refers to a group of neurogenic speech disorders characterized by abnormalities in the strength, speed, range, steadiness, tone, or accuracy of movements required for breathing, phonatory, resonatory, articulatory, or prosodic aspects of speech production. Proper evaluation of speech intelligibility is a key diagnostic step in identifying the effect of treatment on the patients. The proposed system has been developed so as to make this evaluation process simple for both the patient and the clinician. The system automatically estimates the intelligibility score on a scale of 0 to 100 along the lines of Assessment of Intelligibility of Dysarthric Speech (AIDS) score.**

Keywords–Dysarthria; Automatic Intelligibility Assessment; AIDS Score.

I. INTRODUCTION

Dysarthria is a motor speech disorder where permanent brain and/or nerve damage impacts speech-related muscles. These muscles either go limp and loose or become tight and rigid, thereby causing difficulty in speech production. Dysarthria is a common symptom in neurological disorders such as Parkinson's Disease (PD), Huntingtons Disease (HD), Amyotropic Lateral Scleoris (ALS), cerebral palsy or neurological trauma. Additionally, dysarthria may also arise after a traumatic head injury or a side effect of brain tumor. Reduction in intelligibility, audibility, naturalness, and efficiency of vocal communication are the major manifestations observed in dysarthric subjects which leads to slurred speech, hoarse and choppy sound, hypernasal voice and articulation errors [1] [2].

In order to analyze the patients progress and the effects of speech therapy and medication, accurate and consistent assessment of speech intelligibility at regular intervals is important. Traditionally, this assessment has been performed by a trained Speech Language Pathologists (SLP) who use different measures like Hoehn and Yahr scale, AIDS scale, etc. However, due to inter-listener differences, these methods are susceptible to errors and, therefore the development of a standardized method for intelligibility estimation is important [3].

In order to overcome these shortcomings in perceptual assessment, various automatic intelligibility assessment systems have been proposed in the past. In [4], the authors tackled the problem by using an i-vector based approach. A system based on audio descriptors was described in [5], where the features traditionally used to define timbre of musical instruments was modified to address the intelligibility assessment. More recently, spectral subspace analysis has been proven effective for such an assessment in [6]. An in-depth overview of objective assessment techniques for dysarthria intelligibility assessment has been addressed in [7]. It should also be noted that intelligibility assessment is a precursor to recognition of dysarthric speech [8].

A computer based system has the capability of being operated on recorded voice and does not require the patient to physically visit the clinician for intelligibility assessment. With the help of such a system, the patients progress can be electronically stored, analyzed remotely without the patient having to travel to the clinic. In this paper, we propose a novel system for automatically estimating the intelligibility scores from a few utterances of dysarthric speech on a scale of 0 to 100 along the lines of AIDS score. Our main aim is to develop a usable system, which requires minimum effort from the patient and the outcome of the system is easily interpretable by the clinician. This not only decreases the discomfort caused to the patient, but also reduces the errors caused by the previous listeners experience, which is the major drawback of perceptual evaluation.

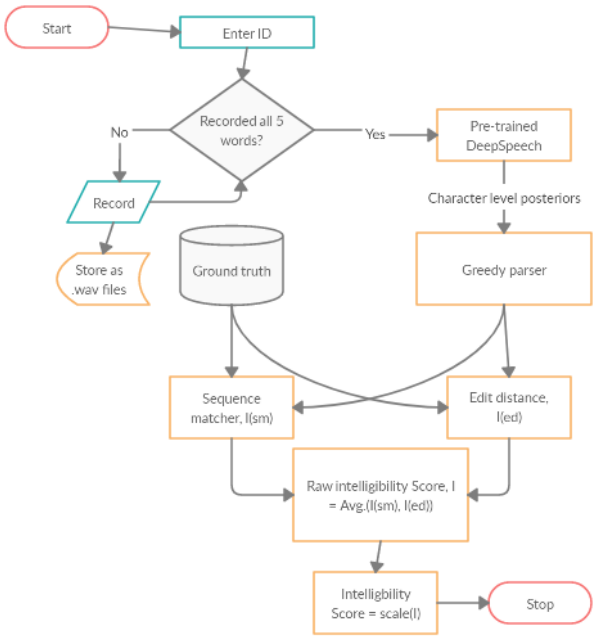

Figure 1. System flowchart.

The rest of the paper is organized as follows. In Section II, we describe the system in detail. In Section III, we present the implementation of the system and we conclude in Section IV.

II. SYSTEM OVERVIEW

The proposed system, in its final form, is very simple to use; we ask the patient to speak and record a set of *five words* which have been identified experimentally from a dataset of 455 words generally chosen for the task of intelligibility assessment [9]. Subsequently, the audio recording of the five words is analyzed to determine the intelligibility score of the patient. Figure 1 shows the complete functionality of the proposed system. From the usage perspective, the patient needs to enter a unique identity number assigned to them to start the assessment process. An audio recorder (sampling at 16 kHz) active for 3 seconds is enabled. This audio recording process is repeated for all the five words and processed by an end to end DeepSpeech (speech-to-character) engine resulting in a string of letters as output (say, o). This string of letters is then compared with the letters corresponding to the original word (say, r) using two different metrics, namely, Levenshtein distance ($l(ed)$) and Sequence Matcher ($l(sm)$) to obtain a raw intelligibility score. The raw score is

 ISBN: 978-1-61208-760-3

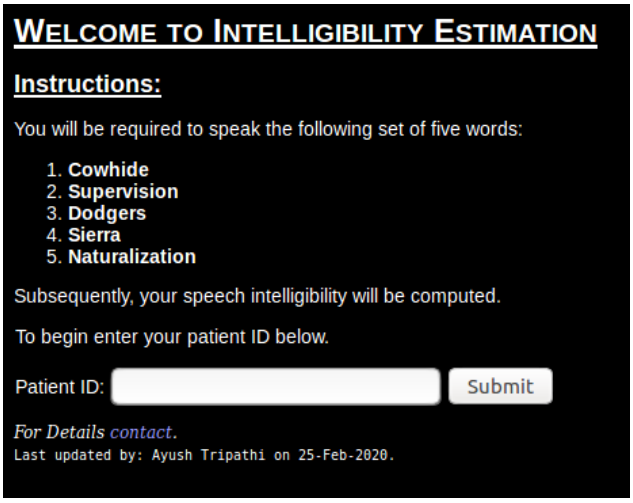

Figure 2. Assessment System Welcome Screen with instructions (Web based).

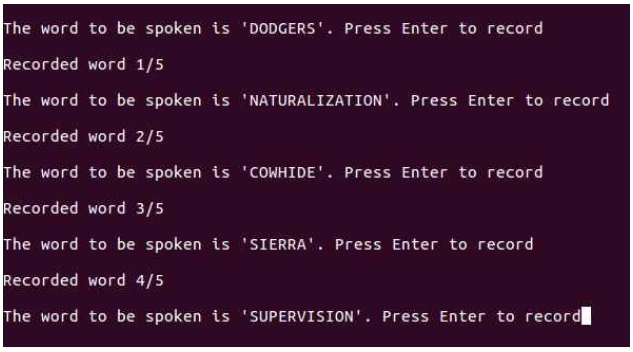

Figure 3. Audio Recording Process.

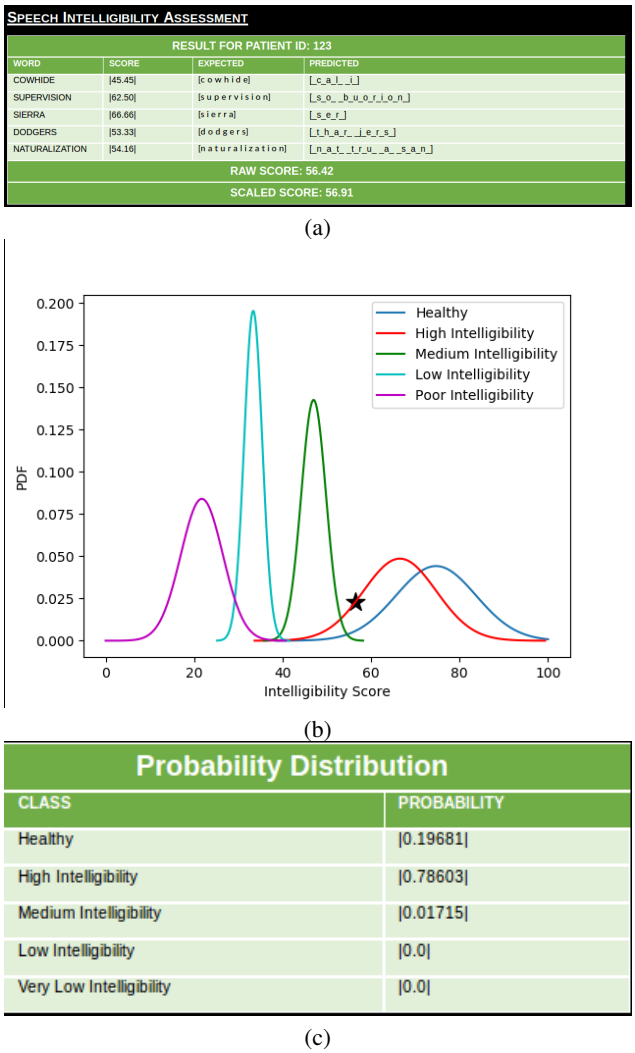

Figure 4. Web based Dysarthria Intelligibility Assessment System.

then normalized against a set of predefined baseline scores to obtain a scaled score. This scaled score is in the range 0 to 100 and is easily interpretable by the clinician who is familiar with the AIDS scale.

III. IMPLEMENTATION

The proposed system is implemented in Python, a general-purpose high level programming language. The system can be broadly divided into three major components, namely, (a) audio recorder, (b) speech-to-character engine and (c) intelligibility score estimator. Individual components are described in detail. The functional system welcome screen with instruction can be seen in Figure 2.

Audio Recorder: The main motivation of the proposed system is to estimate the speech intelligibility of a dysarthric patient by causing minimum discomfort, in terms of the number of words the patient needs to speak. For this purpose, we carefully choose a set of five words namely, DODGERS, NATURALIZATION, COWHIDE, SIERRA and SUPERVISION through extensive experimentation [9]. Notice that the articulation of these words involves a precise control of the *velum* and hence, patients with dysarthria have difficulty in producing these words. These words can be recorded by the patient by using a microphone attached to a desktop (a command line interface to record audio is shown in Figure 3) or can be recorded using any other audio recording utility and sent to the clinician.

Speech-to-character engine: The recorded audio samples are processed using DeepSpeech [10] [11], an open source speech-to-character engine to obtain a string of letters recognized by the Automatic Speech Recognition (ASR) process. Unlike in general purpose ASR, a language model is *not* used for decoding to retain the actual pronunciation of the patient. This pronunciation (or the lack of it) is crucial for identifying intelligibility errors.

Intelligibility Score Estimation: The obtained string of letters (o) is compared to the string of letters (r) corresponding to the original word by using two independent metrics. For example, r is `[c o w h i d e]` and o is `[c a l i]` as seen in Figure 4(a). The first metric is the *Levenshtein distance*, which computes the cost of converting a string o to the reference string r. The second metric is the *Sequence Matcher*, which gives a measure of the similarity between the two strings o and r. In order to find the longest continuous matching subsequence in o and r, sequence matcher is applied recursively to the sequences to the left and to the right of the matching subsequence. Similarity scores obtained from both the metrics are averaged to obtain a raw intelligibility score (see Figure 1) for the patient. Using the intelligibility scores marked by clinicians of 28 patients from UASpeech corpus [12], and the

 ISBN: 978-1-61208-760-3

intelligibility raw scores determined by our system, we normalize the raw score to determine an intelligibility score between 0 and 100. Figure 4(a) shows both the raw and the scaled intelligibility scores. The 28 subjects are divided into 5 categories, namely, *Healthy*, *High Intelligibility*, *Medium Intelligibility*, *Low Intelligibility* and *Very Low* or *Poor Intelligibility*. We model each of these 5 classes with their mean and variance and can be represented as a Gaussian distribution, see Figure 4(b). The scaled intelligibility score is then used to assign a class among these five classes, with a confidence (or probability) of belonging to that class (see Figure 4(b)). So, for a given intelligibility score, we assign the probability of the patient belonging to each of the five classes. As can be seen in Figure 4(c), the patient belongs to the class *High Intelligibility* with probability or confidence 0.786, to the class *Healthy* with a probability of 0.197 and to class *Medium Intelligibility* with a probability of 0.017, thereby suggesting that the patient has a higher probability of belonging to the class *High Intelligibility*. The ability to graphically see the intelligibility score (Figure 4(b)) is advantageous and can be used by the clinician for getting a clear understanding of the patient's progress.

IV. Conclusion

Clinicians dealing with dysarthric patients are in need of an automatic intelligibility assessment system that is consistent, relate-able and usable. In this paper, we propose a working system that is easy to use by both the patient and the clinician. From the patient perspective, it is minimalistic in the sense that it requires the patient to speak only five words. From the clinician perspective, the intelligibility scores output by our system are very easy to interpret because they are relate-able by them to the well known and widely used AIDS score. The graphical representation of the intelligibility scores is an advantage too.

References

[1] P. C. Doyle et al., "Dysarthric speech: a comparison of computerized speech recognition and listener intelligibility." Journal of Rehabilitation Research & Development, vol. 34(3), July 1997, pp. 309–16.

[2] "Dysarthria," https://www.asha.org/public/speech/disorders/dysarthria/, accessed: 03-March-2020.

[3] R. Kent, "Some limits to the auditory-perceptual assessment of speech and voice disorders," American Journal of Speech Language Pathology, vol. 5, no. 3, 1996, pp. 7–23.

[4] D. Martinez, P. Green, and H. Christensen, "Dysarthria intelligibility assessment in a factor analysis total variability space," in ISCA Interspeech, 2013, pp. 2133–2137.

[5] C. Bhat, B. Vachhani, and S. K. Kopparapu, "Automatic assessment of dysarthria severity level using audio descriptors," in 2017 IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP 2017, New Orleans, LA, USA, March 5-9, 2017. IEEE, 2017, pp. 5070–5074. [Online]. Available: https://doi.org/10.1109/ICASSP.2017.7953122

[6] P. Janbakhshi, I. Kodrasi, and H. Bourlard, "Spectral subspace analysis for automatic assessment of pathological speech intelligibility," in ISCA Interspeech, 2019, pp. 3038–3042.

[7] R. Hummel, "Objective estimation of dysarthric speech intelligibility," Master's thesis, Queen's University, Kingston, Ontario, Canada, September 2011. [Online]. Available: https://qspace. library. queensu. ca/handle/1974/6779

[8] C. Bhat, B. Das, B. Vachhani, and S. K. Kopparapu, "Dysarthric speech recognition using time-delay neural network based denoising autoencoder," in Interspeech 2018, 19th Annual Conference of the International Speech Communication Association, Hyderabad, India, 2-6 September 2018, B. Yegnanarayana, Ed. ISCA, 2018, pp. 451–455. [Online]. Available: https://doi.org/10.21437/Interspeech.2018-1754

[9] A. Tripathi, S. Bhosale, and S. K. Kopparapu, "A novel approach for intelligibility assessment in dysarthric subjects," in International Conference on Acoustics, Speech and Signal Processing, 2020.

[10] A. Y. Hannun et al., "Deep speech: Scaling up end-to-end speech recognition," CoRR, vol. abs/1412. 5567, 2014.

[11] Mozilla, "Deep speech 0.4.0," https://github. com/mozilla/DeepSpeech/releases.

[12] H. Kim et al., "Dysarthric speech database for universal access research," in ISCA Interspeech, 2008, pp. 1741–1744.

 ISBN: 978-1-61208-760-3