Weak Speech Supervision: A case study of Dysarthria Severity Classification

Mirali Purohit[1], *Mihir Parmar*[2], *Maitreya Patel*[1], *Harshit Malaviya*[1], *and Hemant A. Patil*[1]
[1]Speech Research Lab, DA-IICT, Gandhinagar, India.
[2]Arizona State University, Tempe, USA.
E-mail: {purohit_mirali, maitreya_patel, harshit_malaviya, hemant_patil}@daiict.ac.in, mihirparmar@asu.edu

***Abstract*—Machine Learning methodologies are making a remarkable contribution, and yielding state-of-the-art results in different speech domains. With this exceptionally significant achievement, a large amount of labeled data is the largest bottleneck in the deployment of these speech systems. To generate massive data, hand-labeling training data is an intensively laborious task. This is problematic for clinical applications where obtaining such data labeled by speech pathologists is expensive and time-consuming. To overcome these problems, we introduce a new paradigm called *Weak Speech Supervision* (WSS), a first-of-its-kind system that helps users to train state-of-the-art classification models without hand-labeling training data. Users can write labeling functions (i.e., weak rules) to generate weak data from the unlabeled training set. In this paper, we provide the efficiency of this methodology via showing the case study of the severity-based binary classification of dysarthric speech. In WSS, we train a classifier on trusted data (labeled with 100% accuracy) via utilizing the weak data (labeled using weak supervision) to make our classifier model more efficient. Analysis of the proposed methodology is performed on Universal Access (UA) corpus. We got on an average 35.68% and 43.83% relative improvement in terms of accuracy and F1-score w.r.t. baselines, respectively.**

***Index Terms*—Dysarthria, Severity-based Classification, Data Scarcity, Weak Supervision, CNN.**

I. Introduction

Speech production is one of the most complex motor coordination processes of the human brain [1]. Dysarthria is a speech disorder in which people lose the natural way to produce speech; which results in slurred, blurred and less intelligible speech. Articulatory parts which are used during speech production can not work properly due to the neurological diseases such as Cerebral Palsy (CP), Parkinson's disease (PD), etc. which causes dysarthria. Losing the natural way of speaking affects one's life tremendously in terms of conveying messages, expressing emotions, using speech technologies like Intelligent Personal Assistants (IPAs), etc. Dysarthric and natural speech are different from speech production-perception perspective [2], and it is hard for machines to recognize and process impaired speech. For this, it is necessary for the machines to recognize the severity of that speech. Understanding the severity can help significantly in improving dysarthric speech [3], [4].

Severity-based classification of impaired speech is a hard and time-consuming process even for speech pathologists [5]. Despite the unprecedented success of Deep Learning (DL)-based techniques, massive labeled data is often needed for classification problem [6]. In deploying Machine Learning (ML)-based systems, hand-labeling training data is an intensively time-consuming, costly, and laborious task. In addition, the performance of these systems degrades in the presence of noisier labels. To solve these problems, *Weak Supervision* is one of the efficient methods to generate noisy or heuristic labels, and training of the model is performed with these weakly labeled data along with trusted data in order to improve the performance [6]. Moreover, dysarthric speech data collection is difficult [7] and there is a scarcity of labeled data for severity-based classification.

In this study, our main aim is to explore weak supervision methodology to label the dysarthric speech data for binary severity-based classification. To this end, we present the *Weak Speech Supervision* (WSS), the first-of-its-kind system that helps users to train state-of-the-art models without hand-labeling severity-based training data of dysarthric speech. WSS is a general framework for many application domains in speech or speaker classification. However, the scope of this study is focused on the severity-based binary classification of dysarthric speech. In our case, we want to classify dysarthric speech into two classes: 1) Low-to-Mid, 2) Mid-to-High. WSS is fundamentally operates on two key steps: 1) Generation of weak labels for unlabeled data, and 2) Utilize weakly labeled data to train classifier. We conclude that WSS achieved on an average 35.68% and 43.83% relative improvement in accuracy and F1-score w.r.t. baseline. Our contributions are summarized as follow: 1) We propose a principled way to learn with weak speech supervision for speech classification tasks; 2) We demonstrate a novel approach for severity-based binary classification of dysarthric speech with limited labeled data; and 3) We conduct experiments on real-world datasets to demonstrate the effectiveness of WSS for given case study.

II. Related Work

In this Section, we briefly show the past attempts have been made for 1) Severity-based classification of dysarthric speech, and 2) Weak Supervision.

A. Severity-based Classification

There are many subjective and objective methods to measure the severity of dysarthria; however, researchers consider the standard measure, which is given by Speech-Language Pathologists (SLP) [8]. The study reported in [9] shows a detailed analysis of the various methodologies for dysarthria severity assessment. Some of these methods determine the severity by extracting the intelligibility of phonemes, single words, narratives, paragraph, etc. Most of the past attempts have been made, mainly focused on the feature-based techniques, and the acoustic modelling for the recognition or detection of severity [10], [11]. Recently, with the advent of ML and DL, the methodology proposed in [12] using different features set and Artificial Neural Network (ANN) has got state-of-the-art results in severity-based classification. With this, data augmentation technique is also used to solve the data scarcity problem in severity-based classification [13]. The method presented in [14] uses a multi-task learning-based approach and Convolutional Neural Network (CNN)-based classifier. One of the significant works in this field is reported in [15] using spectro-temporal features, and ANN and CNN-based classifiers.

B. *Weak Supervision Learning*

The performance of machine learning systems is reliant on labeled data, and show degradation in performance in the presence of noisy labels [16]. Hence, modeling, correcting, and learning with noisy labeled data has been well-studied in [17], [18]. Recently, people are increasingly using the *weak supervision*: cheaper sources of labels that are noisier or heuristic [6]. One method is to generate noisy labels is to heuristically align records of the external knowledge with the data points [19]–[21]. For labeling the data, crowdsourced labels [22], [23], rules and heuristics [24], [25] are used. Several other attempts have been made to achieve weak supervision [26]–[28]. Recently,the state-of-the-art system Snorkel was proposed by [6] for rapid training data creation with Weak Supervision.

In this work, we propose a novel approach that aims to improve the performance of speech classification systems using the WSS technique and provide a case study of severity-based binary classification.

III. Problem Formulation

Our goal is to classify dysarthric speech based on its severity-level. In this study, it is binary classification problem, where dysarthric speech is classified in two severity-based classes: 1) Low-to-Mid (i.e., 0-50% severity-level), and 2) Mid-to-High (i.e., 51-100% severity-level). Let $\mathcal{X} = \{x_i\}_{i=1}^{n}$ denotes the features of dysarthric speech, and $\mathcal{Y} = \{y_i\}_{i=1}^{n} \subset \{0, 1\}$ denotes the corresponding labels. First, labeled data $\mathcal{D} = \{x_i, y_i\}_{i=1}^{n}$ mapped as:

$$f(x) = \begin{cases} y = 0, & \text{if Severity-level is Low-to-Mid,} \\ y = 1, & \text{if Severity-level is Mid-to-High.} \end{cases}$$

In addition, we have a large amount of unlabeled data. For this unlabeled data, we can generate weak labels using a weak labeling function based on dysarthric speech parameters. Formally, we defined a labeling function $h : \hat{\mathcal{X}} \rightarrow \hat{\mathcal{Y}}$, where $\hat{\mathcal{X}} = \{\hat{x}_i\}_{i=1}^{N}$ denotes the features corresponding to unlabeled dysarthric speech, and $\hat{\mathcal{Y}} = \{\hat{y}_i\}_{i=1}^{N}$ denotes the corresponding generated weak labels. We denote this data as $\hat{\mathcal{D}}$, and often $n << N$. Now, we formally define our problem as:

Problem Statement: Given a small manually annotated dysarthric speech data $\mathcal{D}$ for binary severity-based classification, and weakly labeled large data $\hat{\mathcal{D}}$, learn severity-based classifier, $\mathcal{F} : \mathcal{X} \rightarrow \mathcal{Y}$ which can do efficient classification and generalizes for unseen speakers' data as well.

IV. Proposed Methodology

This methodology is using the concept of weak supervision to train a classifier, and is comprised of two major key steps: 1) weak rule extraction to generate weak data, and 2) training discriminative model as illustrated in Fig. 1.

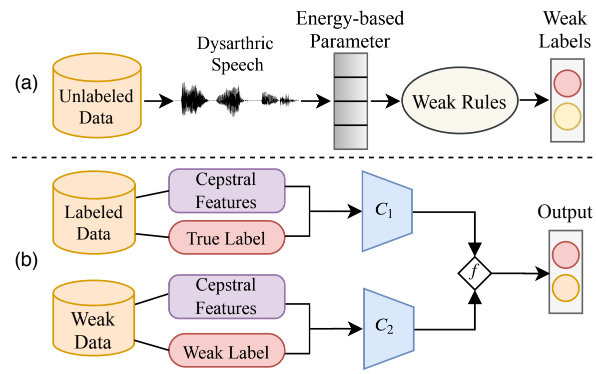


Fig. 1: Schematic representation of (a) weakly labeled data generation using weak rules, and (b) training process of discriminative model using trusted data via utilizing weak data.

A. *Weak Supervision Rules*

The fundamental idea behind this task is to identify labeling functions that generate weak labels for unlabeled data without any manual annotation by humans. Weak rules can be defined by observing patterns, heuristics, some common behavior, external knowledge, etc. [6]. There is a known fact that dysarthria and natural speech are different from speech production-perception perspective [2]. We can observe that dysarthric speech is non-uniform compared to natural speech [29] and hence, these both speech differ in terms of speech rate, percentage of voiced and unvoiced segments, energy, F_0, power, LP residuals, etc. Moreover, dysarthric speech with different severity levels also differ in terms of the above parameters. Hence, we decided to use these parameter-based experiments to define the labeling function. It is observed that as the severity-level changes, the *energy* of dysarthric speech shows significant changes [30]. Hence, the frame-wise *energy* is used as weak rule to define our labeling function.

Energy extraction is done by applying the hamming window of 25 *ms* on the given signal. After that, square of the resulting signal is taken which gives energy vector denoted as $e = [e_1, e_2, ..., e_m]$, where m depends on size of the signal. After that, we calculate the parameter s by taking sum of this energy vector defined as $s = \frac{\sum_{i=1}^{m} e_i}{10^9}$, where 10^9 is a scaling factor. We use this parameter s to define the labeling function. Here, we choose s intuitively via doing experiments using other parameters, such as entropy, mean, standard deviation, etc. We calculate s corresponding to every unlabeled dysarthric speech. Now, we define a function $g:(d,s)\rightarrow y$, where d is dysarthric speech, s is calculated parameter from extracted energy vector, and y is generated label such that,

$$g(d,s)=\begin{cases} y=0, & \text{if s} \geq \phi, \\ y=1, & \text{otherwise,} \end{cases}$$

where ϕ is threshold value define via experiments. Here, we set $\phi=120$ empirically based on the performance on validation data. Specifically, we used the pre-labeled data in order to determine the threshold of weak data generation. ϕ is selected such that generated weak data can be more reliable [6]. We use this $g(d,s)$ to label our training data which generates weak data. This whole process is illustrated in Fig. 1(a).

B. Training Discriminative Model

Our goal is to train a classification model that generalizes beyond the information in the labeling functions and to predict whether the severity of dysarthric speech is Low-to-Mid or Mid-to-High. Here, we devise a new training paradigm that uses weakly labeled data (i.e., weak data) effectively. This training paradigm is illustrated in Fig. 1(b). For this, two different classifiers (i.e., C_1 and C_2) are trained on the labeled data and weak data, respectively. The main idea is to learn the function (f) given in Section III. Suppose that, we have N_1 number of utterances in labeled data set, and N_2 number of utterances in weak data set, where $N_2 > N_1$. Let us assume that the probability predicted by the C_1 and C_2 are $\hat{y}^1$ and $\hat{y}^2$, respectively. Moreover, y^1 and y^2 are the true and weak labels, respectively. Binary Cross-Entropy (BCE) loss is used to optimize the parameters of both the classifiers. We can write objective functions for C_1 and C_2 as follow:

$$\begin{aligned}\mathcal{L}_1&=\frac{1}{N_1}\sum_{i=1}^{N_1}-[y_i^1\log(\hat{y}_i^1)+(1-y_i^1)\log(1-\hat{y}_i^1)],\\ \mathcal{L}_2&=\frac{1}{N_2}\sum_{i=1}^{N_2}-[y_i^2\log(\hat{y}_i^2)+(1-y_i^2)\log(1-\hat{y}_i^2)].\end{aligned} \tag{1}$$

We utilize the training of C_2 on weak data to improve the performance of C_1. To achieve this, we optimize our final objective function as:

$$\mathcal{L}=\mathcal{L}_1+\alpha\mathcal{L}_2 \tag{2}$$

where α is a relative importance of C_2 during training. All the experiments are done for different values of α.

V. Experimental Results

A. Datasets

1) Universal Access Corpus: Universal Access (UA) [31] corpus contains 15 dysarthric speakers (4 females and 11 males) with cerebral palsy, and 13 healthy control speakers (4 females and 9 males). The dataset was created in three blocks, where each block contains 100 uncommon and 155 repeated words. Each speaker produced a total of 765 isolated words, in which 455 words are distinct. The dataset also includes the details of speech intelligibility (in terms of severity-level) of each dysarthric speaker based on transcription task at the word-level, performed by the human listeners.

2) homeService Corpus: The majority of the homeService corpus [32] is recorded in real-world home environments. The corpus contains dysarthric speakers (3 males and 2 females) with cerebral palsy. Each subject's set is composed of 2 subsets: Enrolment Data (ED), and Interaction Data (ID). ED and ID data are recorded using different scenarios [32].

B. Training and Testing Data

For training the classification model, we used data of 8 speakers from UA corpus as trusted data (i.e., labeled data), and 5 speakers from homeService corpus to generate the weak data. From UA, four speakers (F02, F03, M07, and M12) with Low-to-Mid severity-level, and four speakers (F04, F05, M05, and M08) with Mid-to-High severity-level are used as labeled data. To generate the weakly labeled data, we used data of all the speakers (F01, F02, M01, M02, and M03) from homeService as unlabeled data (880 utterances), and annotated this data using WSS. For testing, 55 utterances of each speaker that participated in training from UA are used. In particular, 440 utterances are used to evaluate the performance of the baseline and the proposed method. Moreover, we analyze the proposed method on unseen speakers (i.e., all the remaining speakers of UA corpus whose utterances are not used in training) to show the robustness of proposed method. Training of the proposed method is done for three different scenarios by varying the amount of trusted data: Scenario 1) 300 utterances, 2) 200 utterances, 3) 50 utterances. In every scenario, we used 880 utterances of weak data. Here, the average time of each utterance that used for training and testing is 4.0 seconds.

C. Features and Architectural Details

The training of the classifier is carried out after extracting the Mel Cepstral Coefficients (MCCs) from the speech utterance using AHOCODER [33]. We used 40-dimensional MCCs (including the 0^{th} coefficient) using 25 *ms* Hamming window size, with 5 *ms* frame shift for feature extraction. Convolutional Neural Network (CNN) is used as C_1 and C_2 (shown in Fig. 1b). Baseline systems are also built using CNN. Our network consists of several filter stages, followed by a classification stage. The filter stages include three layers of convolutional filters of kernel size (5×5) with stride 1, and padding 0, followed by max-pooling with kernel size (2×2). The three convolutional layers consist of 16, 32, and 64 filters, respectively. The convolutional layers are followed by 3 fully-connected (FC) layers with dimensions 512, 512, and 1, with the first two layers followed by Rectified Linear Units (ReLU), and last output layer followed by Sigmoid activation function. Batch Normalization is applied before the first FC layer. Stochastic Gradient Descent (SGD) [34] is used as optimizer with learning rate 0.0001. We trained CNN models in traditional way [35].

TABLE I: Analysis of each scenarios over different alpha (α) values and at epoch number 30.

Systems	**Objective Measures**	**Baseline**	**Values of Alpha (α)**												
			0.01	**0.1**	0.2	**0.3**	0.4	0.5	0.6	0.7	0.8	0.9	1	10	50
Scenario 1	Accuracy (%)	72.16	94.49	97.39	97.67	**99.32**	98.12	98.4	98.64	98.58	98.64	98.58	96.81	98.4	98.4
	Precision	0.82	0.95	0.98	0.98	**0.99**	0.98	0.98	0.99	0.99	0.99	0.99	0.97	0.98	0.98
	Recall	0.73	0.95	0.97	0.98	**0.99**	0.98	0.98	0.99	0.99	0.99	0.99	0.97	0.98	0.98
	F1-Score	0.72	0.947	0.974	0.976	**0.993**	0.981	0.984	0.986	0.986	0.986	0.986	0.968	0.984	0.984
Scenario 2	Accuracy (%)	54.26	94.72	97.95	97.27	**98.75**	98.18	96.99	97.5	97.95	97.21	97.27	97.22	97.73	96.36
	Precision	0.76	0.95	0.98	0.97	**0.99**	0.98	0.97	0.97	0.98	0.97	0.97	0.97	0.98	0.97
	Recall	0.55	0.95	0.98	0.97	**0.99**	0.98	0.97	0.97	0.98	0.97	0.97	0.97	0.98	0.96
	F1-Score	0.543	0.947	0.981	0.973	**0.988**	0.982	0.970	0.975	0.979	0.973	0.973	0.973	0.977	0.964
Scenario 3	Accuracy (%)	50	87.04	**97.72**	96.81	93.64	94.09	91.36	92.95	91.59	91.36	91.82	92.95	92.95	89.31
	Precision	0.26	0.90	**0.98**	0.97	0.94	0.95	0.93	0.94	0.93	0.93	0.93	0.94	0.94	0.91
	Recall	0.51	0.87	**0.98**	0.97	0.93	0.94	0.91	0.93	0.92	0.91	0.92	0.93	0.93	0.89
	F1-Score	0.51	0.871	**0.977**	0.968	0.936	0.941	0.914	0.929	0.916	0.914	0.918	0.929	0.929	0.893

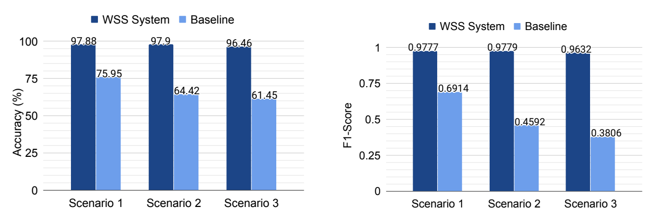
D. Results and Analysis

In this Section, we analyze the effectiveness and robustness of WSS for the severity-based binary classification of dysarthria for three different scenarios. We trained baseline only on trusted data to show the system performance only on labeled data. Accuracy, Precision, Recall, and F1-score are used for objective evaluations, and we provide a detailed analysis of WSS based on these metrics.

1) Analysis of different Scenarios: As shown in Table I, we vary the amount of trusted data during training alongside the α parameter. Here, we keep the amount of the homeService data same while decreasing the amount of the UA corpus step-by-step. For all consecutive scenarios, baseline gets the accuracy of 72.16%, 52.26%, and 50%, and the highest accuracy of the WSS systems are 99.32%, 98.75%, and 97.72%, respectively. In case of each different scenario, we observe that even the smallest value of α gives better results compared to the baseline. Moreover, we can also observe that adding more trusted data, we are getting slightly higher accuracy.

2) Effectiveness of α-parameter: As described in the Section IV-B, α-parameter controls the overall importance of the weakly labeled data while training. For example, $\alpha = 1$ means that weakly labeled data treated with the same importance as trusted data during backpropagation as shown in Eq. 2. In Table I, we have shown results over different values of α for all different scenarios. For scenario 1, the baseline model only gives 72.16% accuracy. However, we can observe that by keeping very small value of α (i.e., 0.01), proposed framework shows the significant improvement in all objective measures. Additionally, we observe the best results for $\alpha = 0.3$ for scenario 1 and scenario 2, and for $\alpha = 0.1$ for scenario 3. However, we increase the α-values, we can see a slight decrease in the results because classifier becomes more biased towards weak data. Interestingly, we can observe that α-values are sensitive towards amount of trusted data in training set. As shown in Table I for scenario 1, accuracy is less fluctuating across different α-values, where trusted data is more in training. As we can see in Table I for scenario 3, this fluctuation in results increases for different α-values, where trusted data is very less in training. Moreover, $\alpha << 1$ degrades the performance of WSS system as well. The reason behind this is that we are nullifying effect of WSS via making importance of weak data almost zero. Hence, we can say that value of α plays a vital role in the performance of WSS system, and need to decide more cautiously when very less trusted data available.

E. Robustness of the WSS system



(a) Accuracy (b) F1-score

Fig. 2: Baseline *vs.* WSS System for all different scenarios for test data of *Unseen* speakers.

To show the robustness of proposed framework, we analyzed the trained WSS model and baseline system on 7 unseen speakers (i.e., speakers' data which are not included while training) for α-value setting of 0.1 for all different scenarios. We test our models on 455 utterances for each speaker, hence, in total of 3185 utterances. It is observed that the WSS system performs similarly while testing on unseen speakers compared to testing on seen speakers, and yielding far better results compared to the baseline in terms of accuracy and F1-score as shown in Fig. 2a and Fig. 2b, respectively.

VI. Summary and Conclusions

In this work, we introduced a novel general framework for speech classification tasks, namely, WSS to overcome the limitation of the scarcity of labeled data. To show the effectiveness of this methodology, we have shown the case study of the severity-based classification of dysarthria. We conclude that WSS significantly improves performance via adding weakly labeled data, and introduced a new training paradigm by utilizing it. Moreover, this work can extend to increase the number of severity classes in dysarthric speech classification, and open the scope by utilizing multi-language dysarthric corpora to produce more weak data. This study provides a scope of improvements in terms of devising different training methods to utilize weak data and extend WSS in different domains of speech classification.

References

[1] RJS Wise, J Greene, C Büchel, and SK Scott, "Brain regions involved in articulation," *The Lancet*, vol. 353, no. 9158, pp. 1057 – 1061, 1999.

[2] Guylaine Le Dorze, Lisa Ouellet, and John Ryalls, "Intonation and speech rate in dysarthric speech," *Journal of communication disorders*, vol. 27, no. 1, pp. 1–18, 1994.

[3] Myung Jong Kim, Joohong Yoo, and Hoirin Kim, "Dysarthric speech recognition using dysarthria-severity-dependent and speaker-adaptive models," in *Interspeech*, 2013, pp. 3622–3626.

[4] Mumtaz Begum Mustafa, Siti Salwah Salim, Noraini Mohamed, Bassam Al-Qatab, and Chng Eng Siong, "Severity-based adaptation with limited data for asr to aid dysarthric speakers," *PloS one*, vol. 9, no. 1, 2014.

[5] Daniel Korzekwa, Roberto Barra-Chicote, Bozena Kostek, Thomas Drugman, and Mateusz Lajszczak, "Interpretable Deep Learning Model for the Detection and Reconstruction of Dysarthric Speech," in *Proc. Interspeech 2019*, 2019, pp. 3890–3894.

[6] Alexander Ratner, Stephen H Bach, Henry Ehrenberg, Jason Fries, Sen Wu, and Christopher Ré, "Snorkel: Rapid training data creation with weak supervision," *Proceedings of the VLDB Endowment*, vol. 11, no. 3, pp. 269–282, 2017.

[7] Ka Ho Wong, *Research and Development of an Articulatory Framework for Dysarthric Speech Analysis*, Ph.D. thesis, Department of Physics, The Chinese University of Hong Kong (Hong Kong), 2017.

[8] Seung Hak Lee, Minje Kim, Han Gil Seo, Byung-Mo Oh, Gangpyo Lee, and Ja-Ho Leigh, "Assessment of dysarthria using one-word speech recognition with hidden markov models," *Journal of Korean medical science*, vol. 34, no. 13, 2019.

[9] Naomi Gurevich and Sydney L Scamihorn, "Speech-language pathologists' use of intelligibility measures in adults with dysarthria," *American journal of speech-language pathology*, vol. 26, no. 3, pp. 873–892, 2017.

[10] Tiago H. Falk, Wai-Yip Chan, and Fraser Shein, "Characterization of atypical vocal source excitation, temporal dynamics and prosody for objective measurement of dysarthric word intelligibility," *Speech Communication*, vol. 54, no. 5, pp. 622 – 631, 2012, Advanced Voice Function Assessment.

[11] Milton Sarria Paja and Tiago H Falk, "Automated dysarthria severity classification for improved objective intelligibility assessment of spastic dysarthric speech," in *Thirteenth Annual Conference of the International Speech Communication Association*, Portland, Oregon, 2012.

[12] 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)*, March 2017, pp. 5070–5074.

[13] Bhavik Vachhani, Chitralekha Bhat, and Sunil Kumar Kopparapu, "Data augmentation using healthy speech for dysarthric speech recognition," in *Proc. Interspeech*, Hyderabad, India, 2018, pp. 471–475.

[14] Juan Camilo Vásquez-Correa, Tomas Arias-Vergara, Juan Rafael Orozco-Arroyave, and Elmar Nöth, "A multitask learning approach to assess the dysarthria severity in patients with parkinson's disease.," in *INTERSPEECH*, Hyderabad, India, 2018, pp. 456–460.

[15] HM Chandrashekar, Veena Karjigi, and N Sreedevi, "Spectro-temporal representation of speech for intelligibility assessment of dysarthria," *IEEE Journal of Selected Topics in Signal Processing*, vol. 14, no. 2, pp. 390–399, 2019.

[16] David F Nettleton, Albert Orriols-Puig, and Albert Fornells, "A study of the effect of different types of noise on the precision of supervised learning techniques," *Artificial Intelligence Review*, , no. 4, pp. 275–306, 2010.

[17] Battista Biggio, Blaine Nelson, and Pavel Laskov, "Support vector machines under adversarial label noise," in *Asian Conference on Machine Learning*, Taoyuan, Taiwan, 2011, pp. 97–112.

[18] Nagarajan Natarajan, Inderjit S Dhillon, Pradeep K Ravikumar, and Ambuj Tewari, "Learning with noisy labels," in *Advances in Neural Information Processing Systems (NIPS)*, Nevada, United States, 2013, pp. 1196–1204.

[19] Razvan Bunescu and Raymond Mooney, "Learning to extract relations from the web using minimal supervision," in *Proceedings of the 45th Annual Meeting of the Association of Computational Linguistics*, Stroudsburg, PA, United States, 2007, pp. 576–583.

[20] Enrique Alfonseca, Katja Filippova, Jean-Yves Delort, and Guillermo Garrido, "Pattern learning for relation extraction with a hierarchical topic model," in *Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, 2012, pp. 54–59.

[21] Mike Mintz, Steven Bills, Rion Snow, and Dan Jurafsky, "Distant supervision for relation extraction without labeled data," in *Proceedings of the Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the AFNLP: Volume 2-Volume 2*, Singapore, 2009, Association for Computational Linguistics, pp. 1003–1011.

[22] Alexander J Quinn and Benjamin B Bederson, "Human computation: a survey and taxonomy of a growing field," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, New York, United States, 2011, ACM, pp. 1403–1412.

[23] Man-Ching Yuen, Irwin King, and Kwong-Sak Leung, "A survey of crowdsourcing systems," in *2011 IEEE Third International Conference on Privacy, Security, Risk and Trust and 2011 IEEE Third International Conference on Social Computing*, 2011, pp. 766–773.

[24] Christopher De Sa, Alex Ratner, Christopher Ré, Jaeho Shin, Feiran Wang, Sen Wu, and Ce Zhang, "Deepdive: Declarative knowledge base construction," *ACM SIGMOD Record*, vol. 45, no. 1, pp. 60–67, 2016.

[25] Theodoros Rekatsinas, Xu Chu, Ihab F Ilyas, and Christopher Ré, "Holoclean: Holistic data repairs with probabilistic inference," *Proceedings of the VLDB Endowment*, vol. 10, no. 11, pp. 1190–1201, 2017.

[26] Percy Liang, Michael I Jordan, and Dan Klein, "Learning from measurements in exponential families," in *Proceedings of the 26th annual International Conference on Machine Learning*, Montreal, Canada, 2009, ACM, pp. 641–648.

[27] Gideon S Mann and Andrew McCallum, "Generalized expectation criteria for semi-supervised learning with weakly labeled data," *Journal of machine learning research*, vol. 11, no. Feb, pp. 955–984, 2010.

[28] Sebastian Riedel, Limin Yao, and Andrew McCallum, "Modeling relations and their mentions without labeled text," in *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*. Springer, 2010, pp. 148–163.

[29] Sanjana Shellikeri, Yana Yunusova, Danielle Thomas, Jordan R Green, and Lorne Zinman, "Compensatory articulation in amyotrophic lateral sclerosis: Tongue and jaw in speech," in *Proceedings of Meetings on Acoustics ICA2013*. Acoustical Society of America, 2013, vol. 19, p. 060061.

[30] Ingo R Titze and Daniel W Martin, "Principles of voice production," 1998.

[31] Heejin Kim, Mark Hasegawa-Johnson, Adrienne Perlman, Jon Gunderson, Thomas S Huang, Kenneth Watkin, and Simone Frame, "Dysarthric speech database for universal access research," in *9th Annual Conference of the International Speech Communication Association*, Brisbane, Australia, 2008.

[32] Mauro Nicolao, Heidi Christensen, Stuart Cunningham, Phil Green, and Thomas Hain, "A framework for collecting realistic recordings of dysarthric speech-the homeservice corpus," in *Proceedings of LREC 2016*. European Language Resources Association, 2016.

[33] D. Erro, I. Sainz, E. Navas, and I. Hernáez, "Improved HNM-based vocoder for statistical synthesizers," in *INTERSPEECH*, Florence, Italy, 2011, pp. 1809–1812.

[34] Sebastian Ruder, "An overview of gradient descent optimization algorithms," *ArXiv*, vol. abs/1609.04747, 2016, {Last Accessed: June 15, 2017}.

[35] Tianyi Liu, Shuangsang Fang, Yuehui Zhao, Peng Wang, and Jun Zhang, "Implementation of training convolutional neural networks," *ArXiv*, vol. abs/1506.01195, 2015, {Last Accessed: June 4, 2015}.