



HAL
open science

Evaluation of a Phone-Based Anomaly Detection Approach for Dysarthric Speech

Imed Laaridh, Corinne Fredouille, Christine Meunier

► **To cite this version:**

Imed Laaridh, Corinne Fredouille, Christine Meunier. Evaluation of a Phone-Based Anomaly Detection Approach for Dysarthric Speech. Interspeech 2016, Sep 2016, San Francisco, United States. pp.223-227, 10.21437/Interspeech.2016-1077 . hal-02102786

HAL Id: hal-02102786

<https://hal.science/hal-02102786>

Submitted on 26 Apr 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Evaluation of a Phone-Based Anomaly Detection Approach for Dysarthric Speech

Imed Laaridh^{1,2}, Corinne Fredouille¹, Christine Meunier²

¹University of Avignon, CERI/LIA, France

²Aix Marseille Université, CNRS, LPL UMR 7309, 13100, Aix-en-Provence, France

imed.laaridh@alumni.univ-avignon.fr, corinne.fredouille@univ-avignon.fr,
christine.meunier@lpl-aix.fr

Abstract

Perceptual evaluation is still the most common method in clinical practice for the diagnosing and the following of the condition progression of people with speech disorders. Many automatic approaches were proposed to provide objective tools to deal with speech disorders and help professionals in the severity evaluation of speech impairments. This paper investigates an automatic phone-based anomaly detection approach implying an automatic text-constrained phone alignment. Here, anomalies are related to speech segments, for which an unexpected acoustic pattern is observed, compared with a normal speech production. This objective tool is applied to French dysarthric speech recordings produced by patients suffering from four different pathologies. The behavior of the anomaly detection approach is studied according to the precision of the automatic phone alignment. Faced with the difficulties of having a gold standard reference, especially for the phone-based anomaly annotation, this behavior is observed on both annotated and non-annotated corpora. As expected, alignment errors (large shifts compared with a manual segmentation) lead to a large amount of anomalies automatically detected. However, about 50% of correctly detected anomalies are not related to alignment errors. This behavior shows that the automatic approach is able to catch irregular acoustic patterns of phones.

Index Terms: Dysarthria, speech disorders, automatic speech processing, objective anomaly detection

1. Introduction

Dysarthria is a motor speech disorder resulting from neurological damages located either in the central or in the peripheral nervous system. This may lead to disturbances in any of the components involved in the speech production, including respiratory, phonatory, resonatory, articulatory and prosodic elements. Consequently, this may be reflected by weakness, spasticity, incoordination, involuntary movements, or abnormal muscle tone, depending on the location of the neurological damage. Dysarthric speech has been studied according to different axes : perceptual evaluation of speech alterations for dysarthria classification [1, 2, 3], perceptual measurement of dysarthria severity, notably related to the speaker's intelligibility [4, 5, 6, 7] or articulatory or/and acoustic analysis [8, 9, 10, 11, 12] in order to observe and characterize the effects of dysarthria in the speech signal. These studies aim at helping clinicians in their knowledge of the speech impairment and its clinical evaluation, crucial for following the condition progression of patients in the case of treatment or/and of speech rehabilitation to enhance them. In this context, automatic speech processing approaches have been seen, very early, as potential solutions to provide objective tools

to deal with speech disorders [13, 14, 15, 16, 17, 18]. In addition, they can also help people with speech disorders in their everyday life through Alternative and Augmented Communication (AAC) tools, involving automatic speech recognition for instance [19, 20, 21, 22, 23].

In the literature, the set of acoustic-perceptual cues including imprecision of consonants, distortion of vowels, slow rate, monopitch, hypernasality is commonly used to characterize the main disturbances of the various types of dysarthria in the speech production. But, more descriptive acoustic and phonetic analyses are still necessary to take into account the large variability in terms of speech alterations observed among people in different groups of diseases and also within the same group [24]. Moreover, so that such analyses are relevant, they require a large number of people with speech disorders, a variety of diseases related to the different types of dysarthria (spastic, flaccid, ataxic, hyper- or hypokinetic, unilateral upper motor neuron, or mixed), various levels of condition progression and of severity degree in order to observe their effects on the speech production, but also possible compensation strategies set up by speakers. Still here, automatic speech processing approaches would be of major interest in the task of focusing the attention of human experts on specific speech segments (among a large amount of speech productions) exhibiting unexpected acoustic patterns compared with a normal speech production.

As reported in [25] in a more general context, anomaly detection refers to the problem of finding patterns in data that do not conform to an expected behavior. In dysarthric speech, anomalies can refer to the unexpected acoustic patterns mentioned above and observed at different units of speech like phones for instance. In previous works [26, 27], the authors have proposed an automatic phone-based anomaly detection approach in two steps : a text-constrained phone alignment to obtain the phone segmentation and a classification of speech segments into normal and abnormal phones (anomalies). The ability of this approach for detecting phone-based anomalies was measured globally, without considering the precision of the phone-based alignment. In [28], the authors studied the large inter-pathology variability between phonetic classes in the text-constrained phone alignment process. In this paper, the authors investigate the impact of the phone-based alignment on the classification process. Indeed, the major hypothesis, which can be raised on the behavior of the detection approach is that an alignment error on a given phone should induce the detection of an anomaly. But, it is interesting to study if this behavior is the only one the detection approach can have or if it is also able to catch distorted acoustic patterns on correctly aligned phones. This study is conducted in the context of dysarthric speech produced by people suffering from four different pathologies.

Faced with the difficulties of having a gold standard reference, especially for the manual annotation of acoustic anomalies at the phone level, the behavior of the detection approach is observed on both annotated and non-annotated corpora.

2. Automatic anomaly detection

The anomaly detection approach studied here relies on two steps. The first step is a text-constrained phone alignment. The second step consists of a two class (normal and abnormal phones) supervised classification. In each class, phones are characterized by a set of features considered as relevant for the discrimination task.

2.1. Automatic phone-based alignment

The segmentation of speech utterances into phones is carried out thanks to an automatic text-constrained phone alignment tool. This tool takes as input the sequence of words pronounced in each utterance and a phonetized phonologically-varied lexicon of words based on a set of 37 French phones. The sequence of words comes from a manual orthographic transcription performed by a human listener, following some specific rules to denote deletions, substitutions, insertions and repetitions of some words/phone sequences. For this manual transcription, inter-pausal units (IPUs) are annotated by the human listener. An IPU is defined as a pause-free unit of speech separated from another IPU by at least 250ms of silence or non-speech. The automatic alignment process is then based on a Viterbi decoding and graph-search algorithms, the core of which is the acoustic modeling of each phone, based on a Hidden Markov Model (HMM). In this work, each phone is modeled using a 3-state context-independent HMM topology which are built using the Maximum Likelihood Estimate paradigm on the basis of about 200 hours of French radiophonic speech recordings [29]. This automatic alignment process results in a couple of start and end boundaries per phone produced in the speech recordings.

2.2. Normal and abnormal speech classification

First, this step aims at characterizing each phone with a set of features found to be relevant for the anomaly detection task. The set of features used is mainly derived from the automatic text-constrained phone alignment outputs. The list of features extracted includes acoustic scores and length (expressed in number of 10ms frames) [26]. The classification task is based on Support Vector Machines (SVM), which have been largely applied to pattern recognition problems [30, 31]. Here, the SVM theory is applied to a two-class problem: discriminating between normal and abnormal phones (anomalies). In order to better take into account the specificities of each phonetic category and to refine abnormal and normal classes, different SVM models are trained by distinguishing the speech productions by gender and phonetic categories (unvoiced consonants, voiced consonants, oral vowels, nasal vowels). The different SVM models are trained using the SVMlight tool (see [32] for more information).

3. Experimental procedure

3.1. Corpora

The current study is based on two speech corpora. The first corpus (LSD) contains 8 dysarthric speakers and 6 control subjects. The dysarthric speakers had been diagnosed with rare lysoso-

Table 1: Information related to patients of both corpora including the number of speakers, the mean and <Min;Max> values of perceptual Dysarthria Severity Degrees (DSD) and the mean and <Min;Max> values of perceptual speech rates per disease.

Disease	Number of speakers	mean DSD <Min;Max>	mean speech rate <Min;Max>
LSD	8	2.0 <1.5;3.0>	-0.5 <-2.9;2.8>
ALS	12	2.0 <0.9;2.9>	-1.3 <-2.7;1.5>
PD	8	0.8 <0.4;1.4>	0.5 <-0.5;1.7>
CA	8	1.3 <0.8;2.1>	-1.1 <-2.2;0.8>

mal storage diseases (LSD), resulting in a mixed dysarthria, and showed disparities in the Dysarthria Severity Degree (DSD). Each patient recorded 3 to 6 longitudinal records approximately every six months. The second corpus (Typaloc [33]) contains 28 dysarthric speakers and 12 control subjects. Unlike the first corpus in which only LSD patients were recorded, this corpus presents various diseases: Amyotrophic Lateral Sclerosis (ALS)/mixed dysarthria, Parkinson’s Disease (PD)/hypokinetic dysarthria and Cerebellar Ataxia (CA)/ataxic dysarthria distributed over various DSDs. All the speakers from both corpora were asked to read the same text, a French fairytale called “Le cordonnier” (The cobbler), as naturally as possible. Speech recordings of patients were evaluated perceptually by a jury of 11 experts who were asked to rate each patient on perceptual items of speech quality. These items included the DSD rated on a scale of 0 to 3 (0 -no dysarthria, 1 -mild, 2 - moderate, 3 -severe dysarthria) and the speech rate on a scale of -3 to 3 (-3 -very slow, 0 -normal, 3 -extremely fast speech rate) on which this paper is focused. In addition, speech utterances were manually segmented by human experts by making corrections, if necessary, to the automatic phone segmentation boundaries. However, since the expert could encounter difficulties when defining phone boundaries, such non-segmentable phone sequences were not considered in the rest of the study. Table 1 provides information on both speech corpora, including the number of patients, the DSD and the speech rate measures per disease. Finally, all the speech utterances of the LSD corpus were annotated by a human expert in order to identify acoustic anomalies at the phone level. Consequently, this corpus was involved in the estimate of the normal and abnormal models required in the classification step described in section 2.2. This annotation task being drastically time-consuming, such annotations are not available for the Typaloc corpus. The “leave-one-out” cross-validation technique is applied so that the corpus-LSD can be used within both the model estimate and testing step, considering the limited amount of labeled data available, notably in terms of abnormal phones extracted from patients (compared to normal ones provided by the control speakers).

3.2. Evaluation

The anomaly detection approach is evaluated by comparing its phone-based decision with the manual annotation of acoustic anomalies available for LSD corpus only, which remains, despite its limits, the gold standard reference for such a task. This evaluation is carried out by comparing the annotations provided for each phone as well as the two adjacent phones (local context). In this case, if the human expert considers a given phone

as abnormal while the automatic approach detects an anomaly on the previous or next phone and not on the given phone, then a right match is counted as well. This approach aims at supporting a “one-phone delay” in the automatic detection due to small boundary shifts in the automatic phone segmentation, for instance. This comparison permits the computation of two evaluation measures which focus on the automatic anomaly detection task: (1) the abnormal-class-based recall measure, ranging from 0 to 1, named *AbnRecall*, is given by the ratio between the number of zones correctly detected as anomalies by the automatic processing and the number of zones labeled as abnormal in the reference; (2) the abnormal-class-based precision measure, ranging from 0 to 1, named *AbnPrec*, is given by the ratio between the number of phones correctly detected as anomalies by the automatic processing and the total number of anomalies reported by the automatic processing (truly or falsely). These measures focus on the detection of abnormal phones only. They have to be considered as complementary in the automatic detection approach assessment. In addition, the precision of the text-constrained phone alignment step is measured by comparing the automatic phone segmentation outputs with the manual segmentations provided by the human expert. This comparison permits the computation of the Start Shift (*SS*) measure, which is given by the absolute value of the difference between the phone start boundaries from the automatic and manual segmentations.

4. Results and discussion

4.1. Anomaly detection on annotated corpus

This section details and discusses the behavior of the automatic anomaly detection approach according to the precision of the text-constrained phone alignment. Here, only the LSD corpus is concerned, for which phone segmentations and anomaly annotations issued from a manual expertise are available.

Table 2: Performance of the anomaly detection approach applied on LSD patients, expressed in terms of *AbnRecall* and *AbnPrec*.

Dysarthric speakers	<i>AbnRecall</i>	<i>AbnPrec</i>
Male average	0.69	0.57
Female average	0.87	0.64
Average	0.78	0.61

Beforehand, table 2 depicts the global evaluation measures without considering the influence of the text-constrained phone alignment. Results show an average *AbnRecall* and *AbnPrec* values of 0.78 and 0.61 respectively. As reported in previous work, the detection approach tends to be more severe by detecting more anomalies than the human expert.

Table 3 displays the distributions of phones, of phones labeled as anomalies by the automatic approach (automatic anomalies), of phones labeled as anomalies by the human expert (manual anomalies) and of phones correctly detected as anomalies by the automatic approach (true positive) according to the *SS* values. This distribution shows that 71% of the phones are aligned within the acceptable range of *SS* values, usually used in phone segmentation evaluation ($[-20\text{ms}, 20\text{ms}]$ i.e. ± 2 frame interval). This rate is quite satisfactory considering the range of Dysarthria Severity Degrees (DSD) given in Table 1 for this

corpus.

Two different behaviors can be pointed out. The first one considers *SS* values outside the ± 2 frame interval. Here, less than 50% of both automatic (44%) and manual (46%) anomalies are represented. Moreover, the probability for a phone to be labeled as anomaly by the automatic approach increases with extreme *SS* values. This behavior is not so surprising given the nature of the features used to characterize phones (mainly issued from the alignment outputs) involved in the normal and abnormal classification. In a similar way, it is interesting to notice that the probability for a phone to be manually labelled as anomaly also increases for high *SS* values. This observation tends to confirm the hypothesis we mentioned in the introduction : an alignment error on a given phone should induce the automatic detection of an anomaly. To this, a part of the phone alignment errors can result from speech alterations. In this case, the text-constrained phone alignment faces more difficulties delimiting phone boundaries. This results in high *SS* values computed for those phones, leading in turn, the classification process to label these phones as anomalies.

The second behavior seen in table 3 considers *SS* values inside the ± 2 frame interval. Here, more than half of the manual anomalies are represented. For these phones, despite the abnormal acoustic patterns underlined by the expert, their irregular nature does not disturb the alignment process. Observing now the distribution of the true positive, roughly 54% of anomalies correctly detected by the automatic approach are not, therefore, related to alignment errors. This observation is very relevant as it demonstrates the ability of the automatic approach to detect abnormal phones, based uniquely on their acoustic irregularities. This supports the quality of the set of features used to characterize phones involved in the anomaly detection process.

4.2. Anomaly detection on non-annotated corpus

Figure 1 reports the distribution of phones according to *SS* values for the different populations of the TYPALOC corpus. As no manual anomaly annotation is available for this corpus, figure 2 reports only the distribution of phones labeled as anomalies by the automatic approach. For comparison purposes, values computed on corpus-LSD are also reported in both figures.

Observing the phone distribution in figure 1, 85%, 64%, 71%

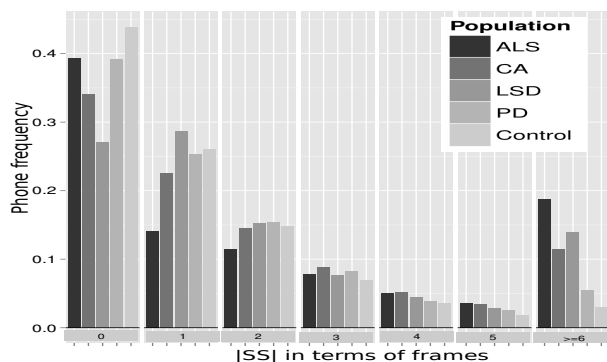


Figure 1: Distribution of phones according to the *SS* values for control, ALS, CA and PD populations (TYPALOC) and for LSD speakers. Each bin refers to a shift of 1 frame (10ms).

and 79% of phones are located in the ± 2 frame interval for control, ALS, CA and PD speakers respectively. If these percentages are not far from the one observed for the LSD patients

Table 3: Distribution of phones, automatic anomalies, manual anomalies and true positives according to the SS values for LSD patients

	$ SS \geq 60\text{ms}$	$ SS =50\text{ms}$	$ SS =40\text{ms}$	$ SS =30\text{ms}$	$ SS =20\text{ms}$	$ SS =10\text{ms}$	$ SS =0\text{ms}$
Phones	1263	259	410	696	1392	2614	2465
Automatic Anomalies	620	92	100	127	199	400	595
Manual Anomalies	363	61	70	81	132	228	310
True Positive	222	32	32	29	59	120	194

Table 4: Relative automatic anomaly rates (%) per pathology and phonetic categories for phones with $SS \in \pm 2$ frames and phones with $SS \notin \pm 2$ frames.

Phonetic Category	$SS \in \pm 2$ frame					$SS \notin \pm 2$ frame				
	Control	CA	PD	ALS	LSD	Control	CA	PD	ALS	LSD
Consonants	7	23	11	32	25	10	35	21	46	40
Vowels	3	10	5	29	12	4	16	13	47	28

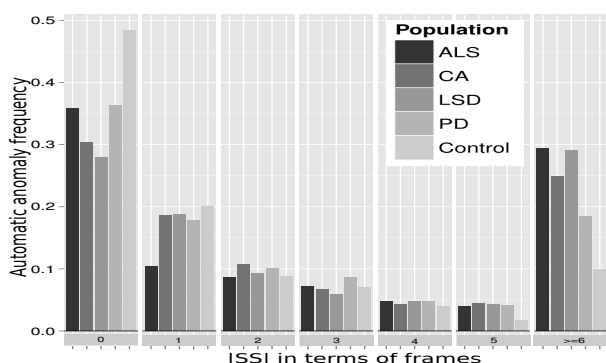


Figure 2: Distribution of automatic anomalies according to the SS values for control, ALS, CA and PD populations (Typaloc) and for LSD speakers. Each bin refers to a shift of 1 frame (10ms).

(71%), the inter-pathology variability could be explained by the disparities in the average DSD and speech rate evaluation across pathologies. Indeed, PD patients related to the lowest DSD average and the fastest speech rate (0.8 and 0.4 respectively) present the best automatic alignment comparing to the ALS patients for whom low speech rate (mean -1.3) and high DSD (mean 2.0) could pose more difficulties to the automatic alignment tool.

Observing the automatic anomalies distribution in figure 2, variability in automatic anomaly values per bin can be observed between diseases. However, it is particularly interesting to point out that the two behaviors previously noticed on the LSD patients can be still observed for all the different diseases, even emphasized notably within the ± 2 frame interval. Indeed, 77%, 55%, 60%, 64% of the anomalies for control, ALS, CA and PD speakers are within this interval, compared to 56% for LSD speakers. The similar trends observed over all the diseases studied here support the assumption that the performance reached by the automatic approach in the task of the anomaly detection over the LSD patients could be transposable to other diseases.

4.3. Automatic alignment and anomaly detection across phonetic categories

Table 4 reports the relative automatic anomaly rates (%) per pathology and phonetic categories for phones with $SS \in \pm 2$

frame interval and for phones with $SS \notin \pm 2$ frames.

First, we observe that consonants exhibit the highest anomaly rates for all populations whatever values of SS observed. This behavior is consistent with dysarthria systemic impact over consonant production.

Overall, for almost all phonetic categories/pathology, higher relative anomaly rates are observed for phones related to $SS \notin \pm 2$ frame interval. This does not mean that their is more anomalies related to alignment errors (outside the ± 2 frames interval) but that the probability of a phone to be found abnormal is higher when the alignment error is more important. This is consistent with the behavior observed in section 4.1 on the annotated corpus.

5. Conclusion

This paper investigates the behavior of an automatic phone-based anomaly detection approach according to the precision of the text-constrained phone alignment involved in the detection process when applied on read French dysarthric speech. This study has shown that the phone-based alignment may be highly sensitive to the presence of phone-based anomalies (resulting in large shifts compared with a manual segmentation), which, in turn, leads to a rather good detection of these anomalies. In addition, 54% of the correctly detected anomalies are not related to alignment errors, which also proves the capability of the automatic approach in detecting unexpected acoustic patterns on well-aligned phones. Further study will investigate how the automatic prediction of some alignment errors may improve the precision of the automatic anomaly detection in terms of *Abn.Prec* or to assign a confidence measure to automatic anomalies.

6. Acknowledgements

This work has been carried out thanks to the support of the BLRI Labex (ANR-11-LABEX-0036) and the A*MIDEX project (ANR-11-IDEX-0001-02) funded by the ‘‘Investissements d’Avenir’’ French government program managed by the ANR, and thanks to the French ANR projet Typaloc (ANR-12-BSH2-0003-03).

7. References

- [1] F. L. Darley, A. E. Aronson, and J. R. Brown, "Differential diagnostic patterns of dysarthria," *Journal of Speech and Hearing Research*, vol. 12, pp. 246–269, 1969.
- [2] J. R. Duffy, *Motor speech disorders: substrates, differential diagnosis and management*. Motsby- Yearbook, St Louis, 2nd edition, 2005.
- [3] F. L. Darley, A. E. Aronson, and J. R. Brown, *Motor speech disorders*. Philadelphia: W. B. Saunders and Co., 1975.
- [4] P. Enderby, "Frenchay dysarthric assessment," *Pro-Ed, Texas*, 1983.
- [5] K. M. Yorkston, E. Strand, and M. Kennedy, "Comprehensibility of dysarthric speech: implications for assessment and treatment planning," *American Journal of Speech Language Pathology*, vol. 55, pp. 55–66, 1996.
- [6] K. C. Hustad, "The relationship between listener comprehension and intelligibility scores for speakers with dysarthria," *Journal of Speech, Language and Hearing Research*, vol. 51(3), pp. 562–573, 2008.
- [7] A. Lowit and R. D. Kent, *Assessment of motor speech disorders*. Plural publishing, 2010, vol. 1.
- [8] R. D. Kent, G. Weismer, J. F. Kent, H. K. Vorperian, and J. R. Duffy, "Acoustic studies of dysarthric speech: Methods, progress, and potential," *The Journal of Communication Disorders*, vol. 32:3, pp. 141–186, 1999.
- [9] M. J. McAuliffe, E. C. Ward, and B. E. Murdoch, "Speech production in parkinson's disease: I. an electropalatographic investigation of tongue-palate contact patterns," *Clinical linguistics & phonetics*, vol. 20, no. 1, pp. 1–18, 2006.
- [10] K. M. Rosen, R. D. Kent, A. L. Delaney, and J. R. Duffy, "Parametric quantitative acoustic analysis of conversation produced by speakers with dysarthria and healthy speakers," *Journal of Speech, Language, Hearing Research*, vol. 49(2), pp. 395–411, 2006.
- [11] J. R. Green, Y. Yunusova, M. S. Kuruvilla, J. Wang, G. L. Pattee, L. Synhorsti, L. Zinman, and J. D. Berry, "Bulbar and speech motor assessment in als: Challenges and future directions," *Amyotrophic Lateral Sclerosis and Frontotemporal Degeneration*, vol. 14(7–8), pp. 494–500, 2013.
- [12] J. A. Whitfield and A. M. Goberman, "Articulatory-acoustic vowel space: Application to clear speech in individuals with parkinson's disease," *Journal of communication disorders*, vol. 51, pp. 19–28, 2014.
- [13] L. J. Ferrier, N. Jarrell, T. Carpenter, and H. C. Shane, "A case study of a dysarthric speaker using the dragondictate voice recognition system," *Journal for Computer Users in Speech and Hearing*, vol. 8(1), pp. 33–52, 1992.
- [14] K. M. Rosen and S. Yampolsky, "Automatic speech recognition and a review of its functioning with dysarthric speech," *Augmentative and Alternative Communication (AAC)*, vol. 16(1), pp. 48–60, 2000.
- [15] J. Carmichael, "Introducing objective acoustic metrics for the Frenchay Dysarthria Assessment procedure," Ph.D. dissertation, University of Sheffield, 2007.
- [16] C. Middag, J.-P. Martens, G. V. Nuffelen, and M. D. Bodt, "Automated intelligibility assessment of pathological speech using phonological features," *EURASIP Journal on Applied Signal Processing*, vol. 2009(1), 2009.
- [17] G. V. Nuffelen, C. Middag, M. D. Bodt, and J.-P. Martens, "Speech technology-based assessment of phoneme intelligibility in dysarthria," *International journal of language and communication disorders*, vol. 44(5), pp. 716–730, 2009.
- [18] M. Kim and H. Kim, "Automatic assessment of dysarthric speech intelligibility based on selected phonetic quality features," in *Computers Helping People with Special Needs*, ser. Lecture Notes in Computer Science, 2012, vol. 7383, pp. 447–450.
- [19] P. Green, J. Carmichael, A. Hatzus, P. Enderby, M. Hawley, and M. Parker, "Automatic speech recognition with sparse training data for dysarthric speakers," in *Proceedings of Interspeech'03*, Geneva, Switzerland, 2003, pp. 1189–1192.
- [20] M. Parker, S. Cunningham, P. Enderby, M. Hawley, and P. Green, "Automatic speech recognition and training for severely dysarthric users of assistive technology: the stardust project," *Clinical Linguistics and Phonetics*, vol. 20(2–3), pp. 149–156, 2006.
- [21] F. Rudzicz, A. K. Namasivayam, and T. Wolff, "The torgo database of acoustic and articulatory speech from speakers with dysarthria," in *Proceedings of the International Conference on Language Resources and Evaluation (LREC'12)*, 2012, pp. 523–541.
- [22] H. Christensen, P. Green, and T. Hain, "Learning speaker-specific pronunciations of disordered speech," in *Proceedings of Interspeech'13*, Lyon, France, 2013.
- [23] H. Christensen, I. Casanueva, S. Cunningham, P. Green, and T. Hain, "homeservice: Voice-enabled assistive technology in the home using cloud-based automatic speech recognition," in *4th Workshop on Speech and Language Processing for Assistive Technologies*, 2013, pp. 29–34.
- [24] B. Tomik and J. Guilloff, "Dysarthria in amyotrophic lateral sclerosis: a review," *Amyotrophic Lateral Sclerosis*, vol. 11 (1–2), pp. 4–15, 2010.
- [25] V. Chandola, A. Banerjee, and V. Kumar, *Anomaly detection: a survey*, University of Minnesota (USA), 2007.
- [26] I. Laaridh, C. Fredouille, and C. Meunier, "Automatic detection of phone-based anomalies in dysarthric speech," *ACM Transactions on accessible computing*, vol. 6, no. 3, pp. 9:1–9:24, May 2015.
- [27] C. Fredouille and G. Pouchoulin, "Automatic detection of abnormal zones in pathological speech," in *Intl Congress of Phonetic Sciences (ICPHS'11)*, Hong Kong, 2011.
- [28] I. Laaridh, C. Fredouille, and C. Meunier, "Automatic speech processing for dysarthria: A study of inter-pathology variability," in *Proceedings of Intl Congress of Phonetic Sciences (ICPHS'15)*, Glasgow, 2015.
- [29] S. Galliano, E. Geoffrois, D. Mostefa, K. Choukri, J.-F. Bonastre, and G. Gravier, "Ester phase ii evaluation campaign for the rich transcription of french broadcast news," in *Proceedings of Interspeech'05*, September 2005, pp. 1149–1152.
- [30] V. Vapnik, *The Nature of Statistical Learning Theory*. New York, NY, USA: Springer-Verlag New York, Inc., 1995.
- [31] B. Scholkopf and A. J. Smola, *Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond*. Cambridge, MA, USA: MIT Press, 2001.
- [32] T. Joachims, "Making large-scale SVM learning practical," in *Advances in Kernel Methods - Support Vector Learning*, B. Schölkopf, C. Burges, and A. Smola, Eds. Cambridge, MA: MIT Press, 1999, ch. 11, pp. 169–184.
- [33] C. Meunier, C. Fougereon, C. Fredouille, B. Bigi, L. Crevier-Buchman, E. Delais-Roussarie, L. Georgeton, A. Ghio, I. Laaridh, T. Legou, C. Pillot-Loiseau, and G. Pouchoulin, "The TYPALOC corpus: A collection of various dysarthric speech recordings in read and spontaneous styles," in *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16)*, May 2016.