



Automatic Prediction of Speech Evaluation Metrics for Dysarthric Speech

Imed Laaridh^{1,2}, Waad Ben Kheder¹, Corinne Fredouille¹, Christine Meunier²

¹University of Avignon, CER/LIA, France

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

{imed.laaridh, waad.ben-kheder, corinne.fredouille}@univ-avignon.fr,
christine.meunier@lpl-aix.fr

Abstract

During the last decades, automatic speech processing systems witnessed an important progress and achieved remarkable reliability. As a result, such technologies have been exploited in new areas and applications including medical practice. In disordered speech evaluation context, perceptual evaluation is still the most common method used in clinical practice for the diagnosing and the following of the condition progression of patients despite its well documented limits (such as subjectivity).

In this paper, we propose an automatic approach for the prediction of dysarthric speech evaluation metrics (intelligibility, severity, articulation impairment) based on the representation of the speech acoustics in the total variability subspace based on the i-vectors paradigm. The proposed approach, evaluated on 129 French dysarthric speakers from the *DesPhoAPady* and *VML* databases, is proven to be efficient for the modeling of patient's production and capable of detecting the evolution of speech quality. Also, low RMSE and high correlation measures are obtained between automatically predicted metrics and perceptual evaluations.

Index Terms: Dysarthria, speech disorders, automatic speech processing, i-vectors, speech intelligibility

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 speech production, including respiratory, phonatory, resonatory, articulatory and prosodic elements. Dysarthric speech has been studied using different approaches: 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] or articulatory or/and acoustic analysis [7, 8, 9] 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. In this context, perceptual evaluation is still the most used method for speech evaluation in clinical practice despite its documented limits such as non-reproducibility and subjectivity.

To cope with these limitations, automatic approaches have been seen, very early, as potential solutions by providing objective tools for intelligibility assessment. In the literature, we can distinguish two main kinds of approaches: those directly based

on automatic speech transcription systems and word transcription error rate as intelligibility score [10, 11], and those which are using automatic speech processing technologies so that relevant information can be extracted from speech and used within an automatic prediction system of the degree of intelligibility [12, 13, 14]. Other automatic approaches focused on a more fine-level evaluation of dysarthric speech such as phones or syllables [15, 16, 17].

In this work, we study the relevance and impact of i-vector representation of dysarthric speech signals for the automatic prediction of different evaluation metrics. The i-vector paradigm is a state-of-the-art approach successfully applied in speaker recognition applications [18]. In [19], this representation, combined with a large set of acoustic, syllable-level, and phonotactic features, was used for the automatic prediction of UPDRS ratings of Parkinson's Disease (PD) patients, in the specific context of Interspeech 2015 ComParE challenge. In [20], the i-vector paradigm was used as a speaker normalization and involved in a more complex classification approach, combining acoustic and articulatory features for the automatic detection of Amyotrophic Lateral Sclerosis (ALS). Finally, in [21, 22], i-vectors were used for the representation of word segments produced by 15 dysarthric speakers resulting in some important correlations between automatically predicted and reference intelligibility measures.

In this paper, the i-vector paradigm is used for the automatic prediction of several dysarthric speech evaluation metrics like intelligibility, severity, and articulation impairment. The proposed approach is applied on 129 dysarthric and control speakers in both gender-dependent and -independent contexts. This work contributes to a larger research project dedicated to the automatic prediction of objective speech evaluation metrics for pathological speech and is not limited to dysarthria. In fact, the assumption is that such methods could be extended to other kinds of speech disorders creating acoustic alterations in the speech signal, such as larynx or head cancers.

2. Motivations

The methodology proposed in this work is similar to the one used in [21]. However, we chose in this work to implement a leave-one-speaker-out strategy where each speaker used as a test subject (for the prediction of the evaluation metric) is not used in the regression model training phase. This choice is motivated by two different reasons. First, the i-vector paradigm has been widely used in speaker recognition context and is proven to well represent and capture speaker acoustics [23]. Thus, the use of the same speakers in the training and testing phases may cause a bias where the automatic regression/prediction process

will no longer predict the speaker’s intelligibility (or any other evaluation metric) but the speaker identity itself. Second, one major application context for this work is longitudinal studies (for scientific research or reeducation purposes). In such a context, the characterization of a speaker’s acoustic traits and not of the quality of his acoustic realization will cause the automatic approach to always predict, for that speaker, similar measures to the one used in the modeling phase with no regard to his condition evolution, and therefore, to the evaluation metric evolution. Finally, the automatic i-vector-based approach is tested here on different corpora, representing different dysarthric classes and pathologies and applied in an automatic prediction task focused on 3 different perceptual evaluation items: speech intelligibility, dysarthria severity degree and articulation impairment.

3. Corpora

The current study is based on several dysarthric speech French corpora, named *VML* and *DesPhoAPady*, described in [24]. In total, 129 dysarthric and control healthy speakers were used:

- 8 dysarthric patients suffering from rare lysosomal storage diseases (LYS) associated with mixed dysarthria;
- 38 participants suffering from ALS associated with mixed dysarthria;
- 31 participants suffering from PD associated with hypokinetic dysarthria;
- 22 participants suffering from Cerebellar Ataxia (CA) associated with ataxic dysarthria;
- 30 control speakers (CTRL).

All participants were asked to read the same text, a fairytale called “*Tic Tac*” (The elves and the shoemaker) containing about 550 phonemes (290 consonants and 260 vowels), as naturally as possible and were recorded inside soundproof or silent rooms during a phoniatry consultation in various medical institutes. All the speech recordings were evaluated perceptually by a jury of 11 experts (10 speech therapist and 1 Ear, Nose and Throat (ENT) specialist/speech pathologist having 7 to 26 years of experience in dysarthric speech perceptual evaluation). The experts were asked to rate all the speakers (including the control speakers) on 9 perceptual items focused on speech quality only (no vocal quality addressed for instance) on a simplified version of the perceptual speech assessment grid provided in the BECD test¹ proposed in [25]. The evaluation was performed on the first minute only of each recording in order to limit the patient’s fatigue effect. These items were rated on quantitative scales of Equal Appearing Interval (EAI) type. This paper focuses on three of them, which are defined as follows:

- the global evaluation of the dysarthria severity degree rated on a scale from 0 to 3 (0=no dysarthria, 1=mild dysarthria, 2=moderate dysarthria, 3=severe dysarthria);
- the global perceptual evaluation of speech intelligibility on a scale from 0 to 3 (0=good intelligibility, 1=medium intelligibility requiring the listener’s attention to understand the message, 2=limited intelligibility requiring a sustained effort, 3=unintelligibility);
- the evaluation of articulation impairment on a scale from 0 to 3 (0=normal to 3=pronounced and constant articulation alterations).

¹BECD is a French acronym which stands for “Batterie d’Évaluation Clinique de la Dysarthrie” (Clinical assessment test for dysarthria)

It is worth noting that the evaluation of speech intelligibility is not based on a word or sentence transcription to provide the intelligibility score like in the Frenchay Dysarthria Assessment test [4], but on the global perception and understanding of the dysarthric speakers’ speech production by the expert jury. Table 1 provides detailed information related to the speakers used in this study including the averaged assessment rates provided by the 11 experts for each perceptual item grouped by gender and population.

Table 1: *Information related to speakers from all corpora grouped by population and gender, including the average values of dysarthria severity, intelligibility and articulation impairment resulting from the 11 experts’ perceptual evaluation*

Disease	# of speakers	dysarthria severity	Intelligibility	Articulation impairment
<i>VML</i>				
LYS - F	4	2.2	1.6	1.9
LYS - M	4	1.9	1.5	1.7
<i>DesPhoAPady</i>				
CA - F	9	1.5	1.1	1.4
PD - F	8	0.6	0.4	0.5
ALS - F	24	1.9	1.3	1.7
CTRL - F	15	0.1	0.0	0.0
CA - M	13	1.5	0.9	1.3
PD - M	23	1.0	0.8	0.8
ALS - M	14	1.9	1.3	1.7
CTRL - M	15	0.2	0.0	0.1

4. Proposed methodology

The proposed approach studied here relies on two steps. The first consists in the parameterization and the representation of each speech utterance in the total variability subspace. Each recording associated with one control or dysarthric speaker will be represented with an i-vector [18].

The second step is a regression from the i-vector subspace to each evaluation metric space (1 dimension). Support Vector Regression (SVR) will be used considering the limited amount of annotated data available for the study. Despite the large number of patients and control speakers available considering the pathological speech context, the amount of data remains limited compared to other “standard” automatic speech processing applications.

4.1. The total variability subspace

The total variability paradigm was first introduced in the context of automatic speaker recognition. In this approach, an i-vector extractor converts a sequence of acoustic vectors into a single low-dimensional vector representing the whole speech utterance. The speaker- and session-dependent super-vector s of concatenated Gaussian Mixture Model (GMM) means is assumed to obey a linear model of the form :

$$s = m + Tw \tag{1}$$

where m is the mean super-vector of the Universal Background Model (UBM), T is the low-rank projection matrix trained using a large dataset by MAP estimation (it represents the “total variability” subspace) and w is a latent variable, called “i-vector”, having a standard normal distribution $\mathcal{N}(0, I)$. The algorithms for the estimation of T and the extraction of i-vectors are described in [26].

4.2. I-vectors extraction

Our experiments operate on 19 LFCC, their 19 first (Δ) and 11 second ($\Delta\Delta$) derivatives. A mean and variance normalization (MVN) is then applied on the LFCC features estimated on the speech portions of each recording, detected using an automatic text-constrained phone alignment. Then, a gender-dependent male 512 component UBM and a total variability matrix T of low rank 400 estimated using French Ester 1&2, REPERE and ETAPE speech corpora (7690 sessions from 2906 speakers)² [27] are used to extract one i-vector per speech recording. The LIA_SpkDet package of the ALIZE open source toolkit [28, 29, 30] is used for the estimation of the total variability matrix and the i-vector extraction.

4.3. Support Vector Regression

In ϵ -SVR, the basic idea is to find a function that has at most ϵ deviation from target reference values for all the training data. When such a task is not feasible, trade-off and slack variables are introduced to cope with the optimization problem [31]. For each test vector, and given the training vectors $x_i \in R^{400}$, $i = 1, \dots, n$, the decision function is:

$$f(x) = \sum_{i=1}^n (\alpha_i - \alpha_i^*) K(x_i, x) + b \quad (2)$$

where α_i and α_i^* are Lagrange multipliers, K is the kernel function and b is the bias. In this work, linear kernels were used.

As reported earlier, a leave-one-speaker-out strategy was employed in all the experiments reported in this work.

5. Results and discussions

To evaluate and study the proposed automatic approach performance on the corpora used in this study, Pearson Correlation (PC), Spearman Correlation (SP) and Root Mean Square Error (RMSE) measures were computed between the automatically predicted and perceptual speech evaluation metrics.

5.1. Gender-independent prediction

The first evaluation context proposed here included all the speech recordings issued from the *VML* and *DesPhoAPady* corpora with no regard to the speakers' gender. This choice was motivated by the intent to use the largest possible amount of data in order to better train and fit the regression model. Table 2 reports the different correlation and RMSE measures for the intelligibility, severity and articulation impairment computed on the different speakers.

First, considering the correlation measures for the different evaluation metrics, we observe that the proposed automatic approach is capable of detecting the evolution of dysarthric speech quality. Indeed, correlation measures higher than 0.8 are observed for the different metrics reaching until 0.88 while predicting the dysarthria severity degree. Also, considering now the RMSE measures, low values are observed for the different evaluation metrics reaching 0.41, 0.44 and 0.44 for intelligibility, severity and articulation impairment respectively. These measures can be considered very low considering the [0,3] interval in which the evaluation takes place. These first results confirm the capacity of the i-vector modeling

²regarding the large number of speakers and sessions required to train a similar UBM model and T matrix for female speakers, these data were not available for this paper.

Table 2: *Pearson Correlation (PC), Spearman Correlation (SC), and Root Mean Square Error (RMSE) measures for the different evaluation metrics - intelligibility, severity and articulation impairment using gender-independent regression (p -value $< 10^{-15}$).*

	Intelligibility	Severity	Articulation
PC	0.83	0.87	0.85
SC	0.84	0.88	0.87
RMSE	0.41	0.44	0.44

in representing the acoustic variability of dysarthric speech and the possible use of such a representation in the prediction of perceptual speech evaluation metrics. Also, this behavior suggests the possible use of such approaches in longitudinal studies and the following of the condition progression of dysarthric speakers in a rehabilitation context.

5.2. Gender-dependent prediction

Considering the acoustic differences between male and female speakers in speech production, we considered using gender-dependent regression models in this section. Table 3 presents the performance of gender-dependent regression models from both the *VML* and *DesPhoAPady* corpora. Performance of the gender-independent model (section 5.1) is also reported per gender for comparison purpose. The leave-one-speaker-out strategy was still implemented for the gender-dependent regression models.

Table 3: *Pearson Correlation (PC), Spearman Correlation (SC), and Root Mean Square Error (RMSE) measures for the different evaluation metrics - intelligibility, severity and articulation impairment for male (M) and female (F) speakers using gender-independent and -independent regressions. (p -value $< 10^{-15}$).*

	Intelligibility	Severity	Articulation
Gender dependent regression			
PC - M	0.78	0.82	0.83
SC - M	0.81	0.81	0.85
RMSE - M	0.44	0.48	0.46
PC - F	0.91	0.93	0.90
SC - F	0.92	0.94	0.90
RMSE - F	0.36	0.39	0.41
Gender independent regression			
PC - M	0.79	0.81	0.81
SC - M	0.78	0.79	0.81
RMSE - M	0.44	0.49	0.46
PC - F	0.87	0.92	0.89
SC - F	0.88	0.94	0.89
RMSE - F	0.39	0.38	0.42

First, comparing the RMSE measures obtained by the gender-dependent and -independent regression models, comparable values are generally reached, except for the intelligibility measures, for which values for the gender-dependent models are slightly better (RMSE measures on female speakers' intelligibility of 0.36 and 0.39 for gender-dependent and -independent models respectively).

This behavior is interesting considering that we use almost only half of the data in the training phase for the gender-dependent models. Such a behavior tends to highlight the importance of the nature of data used for the modeling phase. Indeed, the use of a smaller but more homogeneous training corpus (same speakers' gender) may yield better and more precise prediction models than the use of larger but heterogeneous training corpus.

Second, it can be observed that the automatic prediction over female speakers outperforms systematically the prediction over male speakers for the three speech evaluation items, independently of the regression models used (gender-dependent or -independent model). Regarding the intelligibility prediction task, SC measures of 0.81 and 0.92 as well as RMSE of 0.44 and 0.36 are observed over male and female speakers respectively using the gender-dependent model for instance. This observation is rather unexpected, notably given that the UBM and T matrix used for the i-vector extraction were estimated only on speech recordings issued from a set of male speakers. In-depth analysis is required here to explain this difference of behavior of the prediction approach between male and female speakers.

It is interesting to underline that an oracle regression model, using the same data for the training and test phases, reaches PC measures between 0.99 and 1 and RMSE measures between 0.09 and 0.10 according to the different evaluation items. These results highlight the maximum performance the proposed approach may reach in optimal experimental conditions.

Finally, figures 1 and 2 depict the automatically predicted dysarthria severity and intelligibility measures respectively compared to the reference perceptual evaluation for male and female speakers using gender-dependent regression models. These figures illustrate once again the capacity of the proposed approach in predicting speech evaluation metrics regardless of the speaker's pathology, dysarthric class or severity degree.

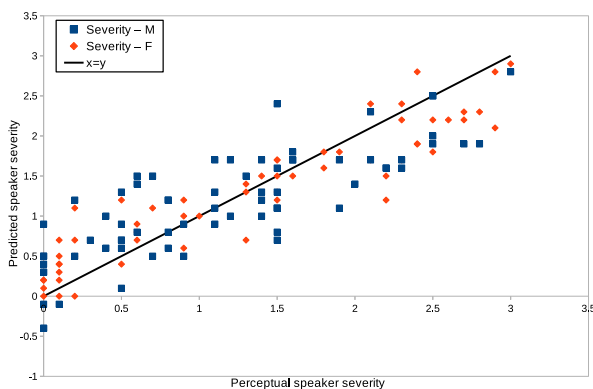


Figure 1: Automatically predicted dysarthria severity degree according to perceptual evaluation and a slope 1 line (black)

6. Conclusions

This paper investigates an automatic approach for the prediction of dysarthric speech evaluation metrics based on the i-vector paradigm and Support Vector Regression-based models. The high correlations and low RMSE measures computed between automatically predicted and perceptual evaluations confirm the capacity of the approach in predicting the intelligibility, severity

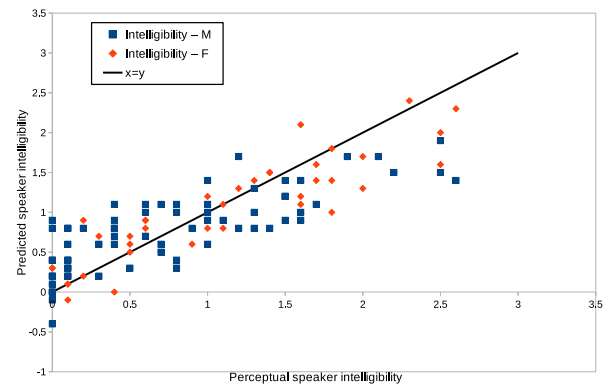


Figure 2: Automatically predicted intelligibility according to perceptual evaluation and a slope 1 line (black)

and articulation impairment measures. This behavior is particularly interesting given the large number of speakers involved in this study as well as the speech impairment diversity in terms of pathology and class of dysarthria.

Future work will investigate the use of this methodology in other pathological speech context, notably for head or larynx cancers. The impact of the phonetic content and/or speech style carried by the speech segments on the prediction task will be also studied. Indeed, the robustness of the proposed approach will be analyzed according to different speech production tasks (isolated words, spontaneous, diadochokinesis of syllables, ...).

7. Acknowledgements

This work has been carried out thanks to the French National Cancer Institute project (INCA - C2SI project). Also, this work was supported by the BLRI Labex (ANR-11-LABEX-0036), the ILCB institute (ANR-16-CONV-0002) and the A*MIDEX project (ANR-11-IDEX-0001-02).

We deeply thank Moez Ajili for his help providing the UBM and the total variability matrix realized within the FABIOLE project (ANR-12-BS03-0011).

8. 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] —, *Motor speech disorders*. Philadelphia: W. B. Saunders and Co., 1975.
- [3] J. R. Duffy, *Motor speech disorders: substrates, differential diagnosis and management*. Motsby- Yearbook, St Louis, 2nd edition, 2005.
- [4] P. Enderby, "Frenchay dysarthric assessment," *Pro-Ed, Texas*, 1983.
- [5] K. C. Hustad, "The relationship between listener comprehension and intelligibility scores for speakers with dysarthria," *Journal of Speech, Language, and Hearing Research*, vol. 51, no. 3, pp. 562–573, 2008.
- [6] A. Lowit and R. D. Kent, *Assessment of motor speech disorders*. Plural publishing, 2010, vol. 1.
- [7] 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.

- [8] J. R. Green, Y. Yunusova, M. S. Kuruville, 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.
- [9] 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.
- [10] P. C. Doyle, H. Leeper, A.-L. Kotler, N. Thomas-Stonell, C. O'Neill, M.-C. Dylke, and K. Rolls, "Dysarthric speech: a comparison of computerized speech recognition and listener intelligibility," *Journal of rehabilitation research and development*, vol. 34(3), pp. 309–316, 1997.
- [11] H. Christensen, S. Cunningham, C. Fox, P. Green, and T. Hain, "A comparative study of adaptive, automatic recognition of disordered speech," in *Proceedings of Interspeech'12*, Portland, USA, 2012.
- [12] C. Middag, J.-P. Martens, G. Van Nuffelen, and M. De Bodt, "Automated intelligibility assessment of pathological speech using phonological features," *EURASIP Journal on Advances in Signal Processing*, vol. 2009, no. 1, pp. 1–9, 2009.
- [13] 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.
- [14] T. Khan, J. Westin, and M. Dougherty, "Classification of speech intelligibility in parkinson's disease," *Biocybernetics and Biomedical Engineering*, vol. 34(1), pp. 35–45, 2014.
- [15] C. Fredouille and G. Pouchoulin, "Automatic detection of abnormal zones in pathological speech," in *Intl Congress of Phonetic Sciences (ICPHs'11)*, Hong Kong, 2011.
- [16] 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.
- [17] —, "Automatic anomaly detection for dysarthria across two speech styles: Read vs spontaneous speech," in *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC 2016)*, Portoro, Slovenia, may 2016.
- [18] N. Dehak, P. J. Kenny, R. Dehak, P. Dumouchel, and P. Ouellet, "Front-end factor analysis for speaker verification," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 19, no. 4, pp. 788–798, 2011.
- [19] G. An, D. G. Brizan, M. Ma, M. Morales, A. R. Syed, and A. Rosenberg, "Automatic recognition of unified parkinsons disease rating from speech with acoustic, i-vector and phonotactic features," in *Proceedings of Interspeech'15*, Dresden, Allemagne, September 2015.
- [20] J. Wang, P. V. Kothalkar, B. Cao, and D. Heitzman, "Towards automatic detection of amyotrophic lateral sclerosis from speech acoustic and articulatory samples," in *Proc. of INTERSPEECH*, 2016.
- [21] D. Martinez, P. Green, and H. C. and, "Dysarthria intelligibility assessment in a factor analysis total variability space," in *Proceedings of Interspeech'13*, Lyon, France, 2013.
- [22] D. Martínez, E. Lleida, P. Green, H. Christensen, A. Ortega, and A. Miguel, "Intelligibility assessment and speech recognizer word accuracy rate prediction for dysarthric speakers in a factor analysis subspace," *ACM Transactions on Accessible Computing (TACCESS)*, vol. 6, no. 3, p. 10, 2015.
- [23] P. Verma and P. K. Das, "i-vectors in speech processing applications: a survey," *International Journal of Speech Technology*, vol. 18, no. 4, pp. 529–546, 2015.
- [24] C. Fougerson, L. Crevier-Buchman, C. Fredouille, A. Ghio, C. Meunier, C. Chevrier-Muller, J.-F. Bonastre, A. Colazo-Simon, C. De-looze, D. Duez, C. Gendrot, T. Legou, N. Lévêque, C. Pillot-Loiseau, S. Pinto, G. Pouchoulin, D. Robert, J. Vaissière, F. Viallet, and C. Vincent, "The despho-apady project: Developing an acoustic-phonetic characterization of dysarthric speech in french," in *Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC'10)*, Valletta, Malta, may 2010.
- [25] P. Auzou and V. Rolland-Monnoury, *Batterie d'évaluation clinique de la dysarthrie*. Édition Ortho, 2006.
- [26] D. Matrouf, N. Scheffer, B. G. Fauve, and J.-F. Bonastre, "A straightforward and efficient implementation of the factor analysis model for speaker verification," in *Interspeech*, 2007, pp. 1242–1245.
- [27] A. Moez, B. Jean-François, B. K. Waad, R. Solange, and K. Juliette, "Phonetic content impact on forensic voice comparison," in *Spoken Language Technology Workshop (SLT), 2016 IEEE. IEEE*, 2016, pp. 210–217.
- [28] J.-F. Bonastre, F. Wils, and S. Meignier, "Alize, a free toolkit for speaker recognition," in *Proceedings of International Conference on Acoustics Speech and Signal Processing (ICASSP'05)*, Philadelphia, USA, 2005.
- [29] J.-F. Bonastre, N. Scheffer, D. Matrouf, C. Fredouille, A. Larcher, A. Preti, G. Pouchoulin, N. W. Evans, B. G. Fauve, and J. S. Mason, "Alize/spkdet: a state-of-the-art open source software for speaker recognition," in *Odyssey*, 2008, p. 20.
- [30] A. Larcher, J.-F. Bonastre, B. G. Fauve, K.-A. Lee, C. Lévy, H. Li, J. S. Mason, and J.-Y. Parfait, "Alize 3.0-open source toolkit for state-of-the-art speaker recognition," in *Interspeech*, 2013, pp. 2768–2772.
- [31] A. J. Smola and B. Schölkopf, "A tutorial on support vector regression," *Statistics and computing*, vol. 14, no. 3, pp. 199–222, 2004.