A MACHINE LEARNING APPROACH TO CLASSIFICATION OF PHONATION MODES IN SINGING

Furkan Yesiler
Universitat Pompeu Fabra
furkan.yesiler@gmail.com

Rafael Ramirez
Universitat Pompeu Fabra
rafael.ramirez@upf.edu

ABSTRACT

Phonation modes are considered as expressive resources in singing voice, and have been defined in four categories (breathy, pressed, neutral and flow) that correspond to different ratios of subglottal pressure and glottal airflow. This work focuses on the automatic classification of phonation modes by analyzing a number of audio descriptors and applying machine learning techniques. The proposed method extends the feature set used in previous works, and uses a correlation-based feature selection algorithm to reduce the feature set dimension. For 10 iterations, cross validation is applied to tune the hyper-parameters of a Multi-Layer Perceptron (MLP) model, and automatic classification is performed on test sets to evaluate the performance. The analysis of the features we propose justifies the decision of extending the feature set. The experiments performed on two reference datasets, separately and combined, result in a mean F-measure of 0.89 for the soprano, 0.97 for the baritone, and 0.93 for the combined datasets. The achieved results outperform the results of previous works.

1. INTRODUCTION

Expression is an essential part of musical performances. Snyder [1] defines expression as a nuance that happens on each occurrence of a musical category, e.g. rhythm and melody patterns. Juslin [2] states that the musical expressions can convey emotions in performances. Expressive resources can be controlled by performers in a number of ways but an inaccurate control over these can seriously decrease the quality of the performance. Musicians frequently use pitch alterations and slight changes in timing to express their style while playing. Loudness, also called intensity or dynamics, and expressive timbre manipulations, the least studied resource compared to the aforementioned three, are closely linked with the emotions the performer wants to express. In this paper, we study phonation modes, an expressive resource of the singing voice.

According to Sundberg [3], phonation modes of the singing voice can be represented as four regions in the two dimensional space spanned by glottal airflow and subglottal pressure. A low glottal airflow combined with a high subglottal pressure results in pressed mode. Breathy mode is the region where there is a high glottal airflow with a low subglottal pressure. While flow mode combines an optimal subglottal pressure and an optimal glottal airflow, neutral mode corresponds to a higher pressure and a lower airflow than flow mode. These different modes help the singers to deliver different expressive and emotional messages to their audiences. Breathy mode can be used to convey sweetness and sexuality while pressed mode delivers a forceful and tense style [4]. Research on phonation modes also may help preventing certain vocal disorders by facilitating the detection of flawed vocal production such as hypo- and/or hyperfunction of the glottis [5].

2. RELATED WORK

As the defining elements of phonation modes, several researches have been conducted to measure the effects of subglottal pressure and transglottal airflow on distinguishing phonation modes. Grillo and Verdolini [6] used physical measurements to prove that the ratio of these two elements can be used for distinguishing pressed, neutral and breathy phonations. Another study, by Millgard et al. [7], found two measures that are correlated with the amount of phonatory pressedness: closing quotient of the glottis and the difference between amplitudes of the first two harmonics in the voice source spectrum.

In order to have information about phonation modes on signal level, a number of features have been studied. Alku et al. [8] proposed the Normalized Amplitude Quotient (NAQ) feature that gives the glottal closing phase to distinguish breathy, pressed and neutral in speaking. Sundberg et al. [9] demonstrated that 73% of the variations in perceived pressedness can be explained by the variations of NAQ. Another feature, Cepstral Peak Prominence (CPP), showed correlation with perceived breathiness [10]. Harmonics-to-Noise Ratio (HNR), Jitter and Shimmer features were investigated by Wakasa et al. [11] to describe their relationship with pressed and neutral modes.

With the curation of the first publicly available dataset for phonation modes in singing by Proutskova et al. [12], there have been several attempts for automatic classification of phonation modes. Proutskova et al. [4] proposed a system to estimate the glottal source waveform to perform automatic classification and the accuracy was in the range of 60% to 75%. A second attempt, by Ioannidis et al. [13], was using Linear Predictive Coding (LPC) related features and achieved a mean F-measure of 0.84 on the same dataset. Their approach was using amplitudes of harmonics, formants, their differences and CPP as features, and a Logistic Model Tree as the classifier. Stoller and Dixon [14] expanded the feature space used in this task with others such as Mel Frequency Cepstral Coefficients (MFCC), and performed an analysis on the behavior of features against phonation modes. They reported that by using Temporal Flatness, separation of the neutral-breathy and pressed flow modes can be demonstrated. CPP and the 0th coefficient of MFCC, in addition to Temporal Flatness, were used to derive a simple rule-based method that resulted in 78% accuracy. Stoller and Dixon performed an automatic classification with various feature sets and using feed-forward neural networks, and achieved a mean F-measure of 0.868. Their results showed that the spectral features, such as MFCC, outperform features from previous works, e.g. NAQ.

In order to extend the scope of phonation modes research, Rouas and Ioannidis [15] have curated another dataset in the same fashion containing recordings of a male baritone singer. They compared two feature sets they described as “Acoustic Descriptors” and “Glottal Features”. The first set includes LPC related features such as CPP and Harmonic-to-Noise Ratio (HNR) while the second includes features, e.g. NAQ, Peakslope and Maximum Dispersion Quotient (MDQ). With using a K-star classifier, they reported that the performance of Acoustic Descriptors was significantly higher than Glottal Features. Combination of both feature sets has increased the results, and on a combined dataset, they reported a mean accuracy of 79%.

In our research, we aim to expand the feature space used for automatic classification to investigate further information about the characteristics of phonation modes. The analysis of the features we propose to use presents relevant information on the task. Before performing an automatic classification, we use feature selection and cross validation to optimize the hyper-parameters of the model, and the obtained model is evaluated on a test set. Our approach on both the soprano and the baritone datasets considerably outperforms previous works.

3. METHODOLOGY

In this section, we outline our methodology which consists of feature extraction, feature selection and training a phonation model using machine learning techniques for performing automatic classification of phonation modes. After introducing the datasets used in our experiments, our decisions on performing each step of this process are explained.

3.1 Datasets

For phonation modes classification, there are 2 published datasets. The first one, by Proutskova et al. [12], contains 909 recordings sung by a female soprano. The pitch range is from A3 to G5, and 9 different vowels were recorded. Not all phonation modes were produced for the entire pitch range, and for some notes, there are multiple recordings. In 2016, Polina Proutskova announced [1] that recordings for flow mode does not represent the definition of Sundberg [3]; thus, they should not be used for phonation mode tasks.

[1] https://osf.io/pa3ha/wiki/home/

Another dataset was curated by Rouas and Ioannidis [15], and contains 487 recordings of single sustained vowels performed by a male baritone. The pitch range of the recordings is from A2 to G4 and 5 different vowels were recorded for each phonation mode.

The experiments in this study are conducted on both datasets and also combined. Due to the information Proutskova provided about the soprano dataset, we consider only breathy, neutral and pressed modes for the combined dataset. To compare the obtained results with previous works, all phonation modes were used to perform automatic classification on the soprano dataset. Since the datasets are not balanced in terms of number of recordings for each phonation mode, balanced versions of the individual and the combined datasets are created. Unless stated otherwise, the balanced versions of datasets are used in our experiments.

3.2 Feature Extraction

In previous research on phonation modes, the performance of various features such as Harmonics, CPP and MFCC are analyzed. We propose, compared to previous works, a wider range of features to analyze the recordings, and utilize a feature selection algorithm to obtain a subset of the features. Table 1 presents the entire list of features used in this study.

	Name	No of Features		Name	No of Features
1	MFCC	52	**11**	H1-H2, H1-A1, H1-A3	3
2	RASTA PLP	52	**12**	CPP	1
3	PLP	52	**13**	Hammarberg Index	1
4	Temporal and Spectral Flatness	3	**14**	Formant Dispersion 1-3	1
5	Spectral Flux	2	**15**	Jitter	1
6	LFCC	32	**16**	Shimmer	1
7	No of Formants	1	**17**	Harmonics-to-Noise Ratio	1
8	Formant 1 Freq., Amp., Bandw.	3	**18**	Energy Below 500Hz	1
9	Formant 2 Freq., Amp., Bandw.	3	**19**	Energy Below 1000Hz	1
10	Formant 3 Freq., Bandw.	2	**20**	Energy Profiles of 500 Hz Wide Bands	15

Table 1. Extracted features

A number of the features from previous works, e.g. MFCC, CPP and harmonics/formant information, are included in our feature set, and the scope of previous works is extended with features such as Perceptual Linear Predictive (PLP) analysis, Relative Spectral Transform of PLP (RASTA PLP) and Linear Frequency Cepstral Coefficients (LFCC). Delta values are calculated for MFCC, PLP and RASTA PLP features.

The reason for expanding the feature set is to gain as much information as possible from the audio data. The first automatic classification study, by Proutskova et al. [4], argued against using spectral features such as MFCC; however, the work of Stoller and Dixon [14] proved that they are useful for the task. PLP, introduced by Hermansky [16], is widely used it speech related tasks e.g. speaker recognition [17] and speech quality analysis [18]. RASTA PLP is a special form of PLP which was introduced by Hermansky et al. [19]. Kepuska and Elharati [20] compared the performances of MFCC, PLP and RASTA PLP features on a speech recognition task, and reported that depending on varying Signal-to-Noise Ratio values, the best performing feature can be different among the 3 features. LFCC is another feature that has been used for speech related tasks such as speaker recognition [21].

For the frame-based features, MFCC, PLP, RASTA PLP, LFCC and Spectral Flux, the mean and the standard deviation values are computed. For MFCC, PLP and RASTA PLP, the mean and the standard deviation values of deltas are included. For representation purposes we refer the mean and the standard deviation of 1st coefficient of PLP, RASTA PLP, MFCC and LFCC as P1M and P1S, R1M and R1S, M1M and M1S, L1M and L1S, respectively. The mean of the 1st delta coefficient of PLP is abbreviated as PD1M. A variety of toolboxes are used for extracting the features: for 1-3 RASTAMAT [22], 4-5 MIR Toolbox [23], 6-10 Praat [24] and 11-20 ProsodyPro with Praat [25]. The definitions and the algorithms used to extract the features can be found in the documentation of the related toolbox.

Stoller and Dixon [14] stated that for extracting MFCC features, using the entire recordings that includes the voice onsets and releases demonstrates a better performance while for the other features, they used trimmed versions of the recordings. In our proposed method, the entire recordings are used for features 1-5 and 7-10, and for features 6, 11-20, the middle 600 ms of the recordings are utilized. Some modifications to the source code of the algorithms are made in terms of facilitating the trimming and labeling.

3.3 Feature Selection

For feature selection and training the machine learning algorithm, we use WEKA toolbox [26] that contains a number of feature selection algorithms such as Wrapper [27] and Information-Gain methods. Correlation-based Feature Selection (CFS) method by Hall [28] is chosen for our experiments.

CFS method, calculates the correlations between the features and their information relevant to classes, and provides a suitable subset. Because of its simplicity, it is computationally efficient and helpful for discarding the redundant features.

3.4 Training the Multi-layer Perceptron

Multi-layer Perceptron (MLP) is one of the most common learning algorithms for classification tasks. In our experiments, we first divide each dataset into two parts in a stratified way: the training subset (90% of the dataset) and the test subset (10% of the dataset). CFS algorithm is used on the training subset to discard the redundant features. For 10 iterations, 10-fold cross validation is performed on the training sets to tune the hyper-parameters of the MLP model. Due to the variety of instances in training sets and the results of feature selection, we test different combinations of hidden layer size, learning rate and momentum coefficient to increase the F-measure resulting from cross validation. For all 3 datasets, we take 0.01 as learning rate, 0.5 as momentum and 2000 as number of epochs. As for hidden layer size, we use one hidden layer with 11 nodes for the soprano dataset, 10 nodes for the baritone dataset and 12 nodes for the combined dataset based on cross validation results.

4. RESULTS

In the following section, we present the results of the experiments for different datasets and feature sets. First, the analysis of PLP, RASTA PLP and LFCC features on the baritone dataset is presented to investigate their relevance to automatic classification of phonation modes. Then, we introduce the feature subset obtained from CFS algorithm for each dataset. Finally, the resulting confusion matrix and the evaluation scores for the each dataset are presented and compared to previous works.

4.1 Feature Analysis

4.1.1 PLP and RASTA PLP

Although RASTA PLP coefficients are derived from PLP, the performances of both feature sets demonstrate distinct behavior against phonation modes. Figure 1 and Figure 2 illustrate the normalized mean and standard deviation of PLP and RASTA PLP features and their delta coefficients. The values are separated by phonation mode and sorted by pitch to observe any resulting pitch dependencies.

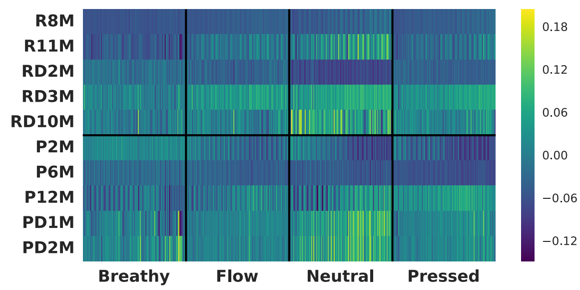

Figure 1. Selected set of mean PLP and RASTA PLP features

For distinguishing neutral and breathy modes from the others, statistics of various PLP and RASTA PLP coefficients demonstrate informative characteristics. Features such as RD2M, RD10M and PD1S can be used to separate neutral mode from the others. For identifying breathy mode, R11M, P2S, P5S and so on, can be considered. While most PLP and RASTA PLP features present relevant information about neutral and breathy modes, P1S differentiates flow mode from the others, P12M flow mode from pressed, and P6M all four modes from each other.

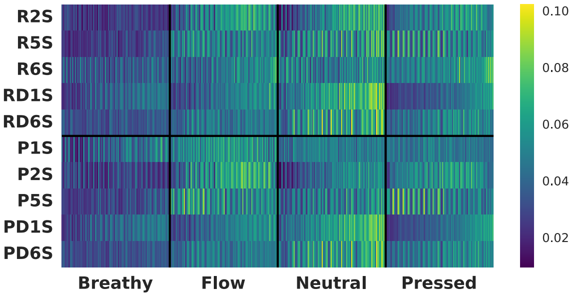

Figure 2. Selected set of standard deviation PLP and RASTA PLP features

In terms of pitch dependency, most standard variation values show correlation with pitch but no trend can be observed in relation to increasing/decreasing number of coefficients.

4.1.2 LFCC

Figure 3, organized in the same way as Figure 1, presents the mean values of LFCCs. As PLP and RASTA PLP, LFCC mostly carry relevant information on separating breathy and neutral modes from the others. L1M, L2M and L9M can be used to differentiate breathy mode from the others, and L12M for detecting neutral mode. For distinguishing pressed mode, L10M presents relevant information, and L2M and L11M outputs observable difference between all four modes. Slight pitch dependencies can be observed for L1M and L12M.

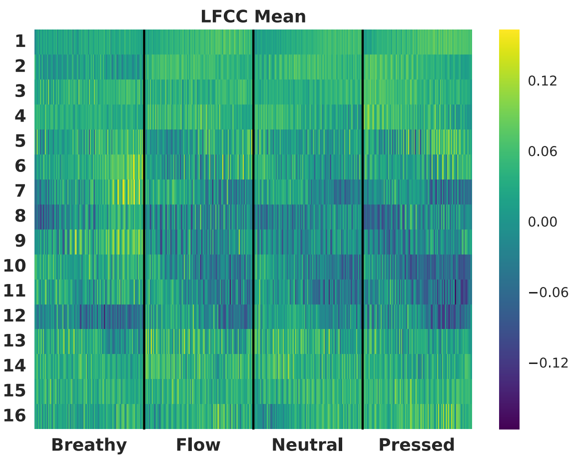

Figure 3. Mean LFCC values by phonation type

4.1.3 Selected Features from the Previous Work

In the works of Proutskova et al. [4], Stoller and Dixon [14], and Rouas and Ioannidis [15], the soprano dataset was used to analyze the behavior of features. Rouas and Ioannidis [15] evaluate the performance of Acoustic and Glottal features on both the soprano and the baritone datasets. In Figure 4, we present distributions (max, min and quartiles) of a selected set of the feature values from previous works. For separating pressed mode, M1M and H1-H2 demonstrate relevant information on the soprano dataset while the same does not apply on the baritone dataset. CPP is useful for separating breathy mode on both datasets but the distributions of neutral and pressed modes differ. HNR presents lower results for breathy mode in the soprano dataset however the same correlation is not observed on the baritone dataset.

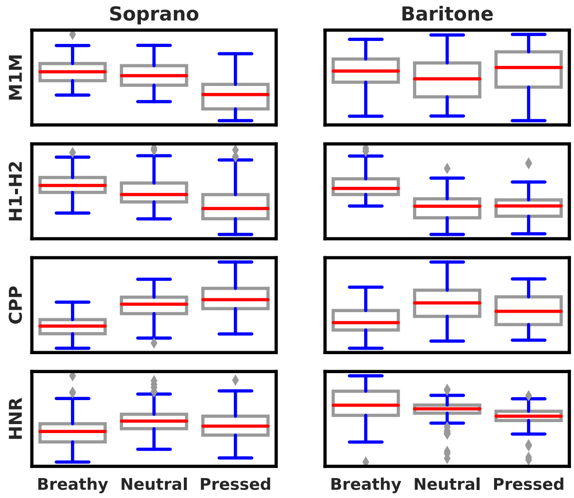

Figure 4. Distributions of selected set of features on soprano and baritone datasets

4.1.4 Selected Features from CFS algorithm

After the feature extraction step, for each individual and the combined dataset, we obtain training subsets and perform CFS algorithm to select the best performing feature subsets. Due to the random nature of creating training subsets, each obtained feature subset differs from the ones in other iterations. Table 2 presents the features that were included in the feature selection results in every iteration for the combined dataset. Although this subset proves its importance on the task, these features should be combined with the other selected features in every iteration to obtain the final scores of the experiments.

No	Features	No	Features
1	MD3S	**9**	L2M
2	MD8S	**10**	L9M
3	MD9S	**11**	L12M
4	R1M	**12**	H1-H2
5	R2S	**13**	H1-A3
6	P8M	**14**	CPP
7	PD7S	**15**	Hammarberg Index
8	Formant 3 Bandwidth	**16**	Jitter

Table 2. The most common features selected by CFS algorithm

4.2 Automatic Classification using MLP

In Table 3, the mean F-measure values for the soprano and the baritone datasets with comparison to previous works are presented. We observe that expanding the feature set and using a feature selection algorithm considerably increased the mean F-measure and accuracy. The obtained mean F-measure for the soprano dataset spans in range of (0.808, 0.961) and for the baritone dataset (0.953, 0.977).

	Dataset	F-measure	Accuracy
Proutskova et al. [4]	Soprano	-	60-75%
Ioannidis et al. [13]	Soprano	0.84	-
Stoller and Dixon [14]	Soprano	0.868	-
Rouas and Ionnidis [15]	Soprano	-	81.62%
	Baritone	-	88.51%
The Proposed Method	Soprano	0.898	89.83%
	Baritone	0.97	96.98%

Table 3. F-measure and Accuracy Comparison

The resulting confusion matrix of our experiments with the combined dataset is shown in Table 4. The mean F-measure is 0.93 , and the values from 0.908 to 0.96 are observed. Due to Proutskova's announcement about flow mode in the soprano dataset, the instances of the combined dataset consist of other 3 phonation modes. The confusion matrix shows that the most misclassified modes are breathy and neutral modes while separating pressed mode from the others demonstrated better results.

	Breathy	Neutral	Pressed
Breathy	236	17	1
Neutral	21	227	4
Pressed	2	8	245

Table 4. Confusion matrix of results on the combined set

5. DISCUSSION

The feature analysis performed on the baritone set presents the behavior of PLP, RASTA PLP and LFCC features on phonation modes. For detecting breathy and neutral modes, various features such as R11M, P5S, L2M and L12M demonstrate relevant information. The distributions of P6M and L11M facilitate separating all phonation modes from each other. The analysis justifies the decision of expanding the feature set.

The analysis regarding the effect of using different datasets on the distributions of a selected set of features chosen from previous works points out the possible problematic conclusions caused by having only 2 datasets to analyze the phonation modes in singing. The potential reasons for the differences (gender, voice type etc.) must be further analyzed.

For the automatic classification step, the combined dataset includes recordings of breathy, neutral and pressed mode from both the soprano and the baritone datasets. The features selected in all 10 iterations show that PLP, RASTA PLP and LFCC support our analysis. Delta values of MFCC and PLP present useful information for the task. Unlike the Information Gain method, the CFS algorithm does not rank the features according to their relevance, thus, the importance of each individual feature in a selected feature subset was not reported, and is subject to further study.

After tuning the hyper-parameters of the MLP model with cross validation on the training subsets, the evaluation results on the test subsets demonstrated considerable improvements compared to previous works. Possible elements of this improvement include expanding the feature set and performing feature selection. For the soprano dataset, the proposed method achieved a mean F-measure of 0.898 in comparison to 0.868 from the work of Stoller and Dixon [14]. The mean accuracy of 96.98% on the baritone dataset outperforms the mean accuracy of 88.51% achieved by Rouas and Ioannidis [15]. The mean F-measure of the proposed method on the combined dataset with breathy, neutral and pressed modes is 0.93. An extended dataset consists of recordings of all phonation modes performed by various singers should be studied in terms of developing a generic classifier.

6. CONCLUSION AND FUTURE WORK

The research presented in this paper explores further analysis of phonation modes in singing based on previous works. Features such as PLP, RASTA PLP and LFCC are investigated for their correlations to automatic classification of phonation modes. An MLP model was used to perform the classification, and an improvement in performance over previous works is observed. The mean F-measure of 0.898 and the mean accuracy of 96.98% was achieved for the soprano dataset and the baritone dataset, respectively. For the combined dataset with breathy, neutral and pressed modes, our method demonstrated a mean F-measure of 0.93.

Our future work regarding the phonation modes includes curating new datasets in order to generalize the analysis on behavior of the features. An application for real time analysis of phonation modes will be created to investigate the effect of automatic classification on the vocal education. Finally, we will design a notation system for phonation modes to be used along with the already existing pitch and duration notations in order to extend the scope of singing voice transcription models.

Acknowledgments

This work was partly sponsored by the Spanish TIN project TIMUL (TIN2013-48152-C2-2-R), and the European Union Horizon 2020 research and innovation program under grant agreement No. 688269 (TELMI project).

7. REFERENCES

[1] R. Snyder, *Music and memory*. MIT Press, 2000.

[2] P. N. Juslin, "Five facets of musical expression: A psychologist's perspective on music performance," in *Psychology of Music*, vol. 31, no. 3, 2003, pp. 273–302.

[3] J. Sundberg, *The Science of the singing voice*. Northern Illinois University Press, 1987.

[4] P. Proutskova, C. Rhodes, T. Crawford, and G. Wiggins, "Breathy, resonant, pressed automatic detection of phonation mode from audio recordings of singing," in *Journal of New Music Research*, vol. 42, no. 2, 2013, pp. 171–186.

[5] E. Froeschels, "Hygiene of the voice," in *Archives of Otolaryngology - Head and Neck Surgery*, vol. 38, no. 2, 1943, pp. 122–130.

[6] E. U. Grillo and K. Verdolini, "Evidence for distinguishing pressed, normal, resonant, and breathy voice qualities by laryngeal resistance and vocal efficiency in vocally trained subjects," in *Journal of Voice*, vol. 22, no. 5, 2008, pp. 546–552.

[7] M. Millgard, T. Fors, and J. Sundberg, "Flow glottogram characteristics and perceived degree of phonatory pressedness," in *Journal of Voice*, vol. 30, no. 3, 2016, pp. 287–292.

[8] P. Alku, T. Backstrom, and E. Vilkman, "Normalized amplitude quotient for parametrization of the glottal flow," in *The Journal of the Acoustical Society of America*, vol. 112, no. 2, 2002, pp. 701–710.

[9] J. Sundberg, M. Thalen, P. Alku, and E. Vilkman, "Estimating perceived phonatory pressedness in singing from flow glottograms," in *Journal of Voice*, vol. 18, no. 1, 2004, pp. 56–62.

[10] J. Hillenbrand, R. A. Cleveland, and R. L. Erickson, "Acoustic correlates of breathy vocal quality," in *Journal of Speech Language and Hearing Research*, vol. 37, no. 4, 1994, pp. 769–778.

[11] K. Wakasa, M. Matsubara, Y. Hiraga, and H. Terasawa, "Acoustic characteristics of pressed and normal phonations in choir singing by male singers," in *Proc. of the 2017 Int. Symposium on Musical Acoustics*, Montreal, 2017, pp. 136–139.

[12] P. Proutskova, C. Rhodes, T. Crawford, and G. Wiggins, "Breathy or resonant - a controlled and curated dataset for phonation mode detection in singing," in *Proc. of the 13th Int. Society for Music Information Retrieval Conf. (ISMIR 2012)*, Porto, 2012, pp. 589–594.

[13] L. Ioannidis, J.-L. Rouas, and M. Desainte-Catherine, "Caractérisation et classification automatique des modes phonatoires en voix chantée," in *XXXèmes Journées d'études sur la parole*, Le Mans, France, 2014.

[14] D. Stoller and S. Dixon, "Analysis and classification of phonation modes in singing," in *Proc. of the 17th Int. Society for Music Information Retrieval Conf. (ISMIR 2016)*, New York City, 2016, pp. 80–86.

[15] J.-L. Rouas and L. Ioannidis, "Automatic classification of phonation modes in singing voice: Towards singing style characterisation and application to ethnomusicological recordings," in *Interspeech 2016*, San Francisco, 2016, pp. 150–154.

[16] H. Hermansky, "Perceptual linear predictive (PLP) analysis of speech," in *The Journal of the Acoustical Society of America*, vol. 87, no. 4, 1990, pp. 1738–1752.

[17] J. Kacur and P. Truchly, "Acoustic and auxiliary speech features for speaker identification system," in *57th Int. Symposium ELMAR (ELMAR)*, Zadar, 2015, pp. 109–112.

[18] H. Salehi and V. Parsa, "Nonintrusive speech quality estimation based on perceptual linear prediction," in *2016 IEEE Canadian Conf. on Electrical and Computer Engineering (CCECE)*, Vancouver, 2016, pp. 1–4.

[19] H. Hermansky, N. Morgan, A. Bayya, and P. Kohn, "Compensation for the effect of the communication channel in auditory-like analysis of speech (rasta-plp)," in *Proc. of Eurospeech '91*, Genova, 1991, pp. 1367–1371.

[20] V. Z. Kepuska and H. A. Elharati, "Robust speech recognition system using conventional and hybrid features of MFCC, LPCC, PLP, RASTA-PLP and hidden markov model classifier in noisy conditions," in *Journal of Computer and Communications*, vol. 03, no. 06, 2015, pp. 1–9.

[21] X. Zhou, D. Garcia-Romero, R. Duraiswami, C. Espy-Wilson, and S. Shamma, "Linear versus mel frequency cepstral coefficients for speaker recognition," in *2011 IEEE Workshop on Automatic Speech Recognition & Understanding*, Waikoloa, 2011, pp. 559–564.

[22] D. P. W. Ellis, "PLP and RASTA (and MFCC, and inversion) in Matlab," 2005, [Accessed 26- May- 2018]. [Online]. Available: http://www.ee.columbia.edu/~dpwe/resources/matlab/rastamat/

[23] O. Lartillot and P. Toiviainen, "A Matlab toolbox for musical feature extraction from audio," in *Proc. of the 10th Int. Conference on Digital Audio Effects*, Bordeaux, 2007.

[24] P. Boersma and D. Weenink, "Praat: doing phonetics by computer [computer program]," 2018, [Accessed 26- May- 2018]. [Online]. Available: http://www.praat.org/

[25] Y. Xu, "ProsodyPro - A tool for large-scale systematic prosody analysis," in *Proc. of Tools and Resources for the Analysis of Speech Prosody (TRASP 2013)*, Aix-en-Provence, 2013, pp. 7–10.

[26] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, and I. H. Witten, "The WEKA data mining software: an update," in *SIGKDD Explorations*, vol. 11, no. 1, 2009, pp. 10–18.

[27] R. Kohavi and G. H. John, "Wrappers for feature subset selection," in *Artificial Intelligence*, vol. 97, no. 1-2, 1997, pp. 273–324.

[28] M. A. Hall, "Correlation-based feature selection for machine learning," Ph.D. dissertation, The University of Waikato, 1999.