Discrimination between pathological voice categories using matching pursuit

Ashwini Jaya Kumar, Khalid Daoudi

Discrimination between pathological voice categories using Matching Pursuit

Ashwini Jaya Kumar
GeoStat team, INRIA
Bordeaux-Sud Ouest, France
Email: ashwini.jaya-kumar@inria.fr

Khalid Daoudi
GeoStat team, INRIA
Bordeaux-Sud Ouest, France
Email: khalid.daoudi@inria.fr

***Abstract*—There are several methods in the literature for pathological voice classification but there are very few methods which can classify pathological sub-groups. An attempt is made here to classify pathological sub-groups using matching pursuit decomposition method and is compared with PRAAT. Random forest classifier is used and frequency band of the atoms are used as feature. The result shows that we can classify adductor spasmodic dysphonia, keratosis and vocal nodules in a class of voices consisting of adductor spasmodic dysphonia, keratosis, paralysis, vocal nodules and vocal fold polyps with reasonably good classification accuracy. Both matching pursuit (MP) and PRAAT shows comparable classification scores but using MP is more advantageous over PRAAT since it doesn't rely on pitch information and extraction of pitch information in a pathological signal is a complex problem.**

I. Introduction

Pathological voice classification is of greater importance with increase in vocal cord disorders and neural disorders which leads to abnormal voice. Early detection of vocal cord disorder may help patients to take remedy before it is worsened. There are number of methods in the literature to classify normal and pathological voices [1], [2], [3], [4]. There is no single method for normal and pathological voice classification which is robust for all the available pathological databases and for real time application. This shows the difficulty in classifying pathological voice from the normal voice. The classification problem becomes more intense if it is within the pathological voices. Pathological voice is a class which contains wave files of several vocal disorders with different severity level, age, gender,..etc. There are not many methods available in the literature which has addressed this problem of pathological voice classification within itself. In [5], classification of vocal fold polyp against keratosis leukoplakia, adductor spasmodic dysphonia and vocal nodules is made using modulation spectra as a feature. In [6], classification of vocal polyp and vocal nodules against keratosis leukoplakia, adductor spasmodic dysphonia is made using adaptive growth of wavelet packet tree, based on the criterion of local discriminant bases. In this paper it is shown that adductor spasmodic dysphonia, keratosis leukoplakia and vocal nodules can be classified in a group consisting of adductor spasmodic dysphonia, keratosis leukoplakia, paralysis, vocal nodules and vocal fold polyp using frequency information of the atoms obtained by matching pursuit algorithm. The centre frequency of the atoms is used to classify normal and pathological voices in [3]. In this paper, depending on the centre frequency of the atoms, it is grouped into 12 bands and the number of atoms falling in each frequency band is used as a feature for classification.

PRAAT is the most popularly used platform for voice analysis, both normal and pathological voices. Since pathological voices are dynamic, turbulent and non-stationary in nature, any pitch detection algorithm which is applicable for normal voices cannot be directly used for pathological voices unless it is validated. It has been validated in [8] of using *Fo* (pitch) estimation algorithm which PRAAT [7] uses on pathological voices and its performance is compared with other methods. In this paper, we are using two approaches for pathological subgroup classification, one is matching pursuit (MP) and the other one is PRAAT. The MP approach has already been proved for normal v/s pathological voice classification in [9]. Although here the result shows that both MP and PRAAT methods show comparable results, MP is more reliable since it doesn't require pitch estimation, which is a complex issue for pathological voices.

In section II, Matching Pursuit (MP) algorithm is explained briefly, feature extraction procedure is explained in section III. Classification results are explained in section IV and finally conclusion is provided in section V.

II. Matching Pursuit Algorithm

The principle of matching pursuit (MP) algorithm is decomposition of a signal $x(t)$ into a linear combination of Time-Frequency (TF) functions (also called atoms) $a_m(t)$ selected from a redundant dictionary of atoms generated by translation, scaling and modulation of complex sinusoids [3].

$$x(t) = \sum_{m=0}^{\infty} \alpha_m a_m(t) \tag{1}$$

where α_m are the expansion coefficients and TF function $a_m(t)$ is defined by,

$$a_m(t) = a(\frac{t - p_m}{s_m}) exp\{j(2\pi f_m t + \phi_m)\} \tag{2}$$

where p_m is the translation parameter, f_m and ϕ_m are the frequency and phase of the exponential function respectively, s_m is the scale factor used to control the width of the window function.

Matching pursuit is a greedy algorithm which iteratively approximates the signal $x(t)$ by projecting it onto the over-complete dictionary D. At each iteration *m*, the MP algorithm

looks for the atom $a_m(t)$ which is the most strongly correlated with the signal $x(t)$ i.e. which has the highest absolute inner product with the signal.

After *M* iterations, $x(t)$ is decomposed as,

$$x(t) = \sum_{m=0}^{M-1} < x_m(t), \hat{a}_m(t) > \hat{a}_m(t) + x_m(t) \quad (3)$$

$$\hat{a}_m(t) = \arg\max_{a \in D} | < x_m(t), a > | \quad (4)$$

where $< ., . >$ is the Hermitian inner product, D is the dictionary containing atoms and $| < x_m(t), a > |$ is the sparse code vector.

Matching Pursuit Tool Kit [10] which efficiently implements the matching pursuit algorithm is used to decompose the signal into atoms using Gabor dictionary in this work.

III. Feature Extraction

A. Feature extraction by Matching Pursuit

A signal is decomposed into atoms using matching pursuit algorithm as described in the above section. The frequency of the obtained set of atoms after *M* iteration is observed and grouped into *N* bands depending on the centre frequency (f_j) of the individual atom. The sampling frequency of the wave files used here is 25kHz and hence according to nyquist theorem, the signal frequency is less than 12.5kHz. The frequency range from 1Hz to 12.5kHz is divided into 12 bands (shown in table I), with 1 kHz frequency range in each band except the Band 12, which has 1.5kHz frequency range. After grouping, the number of atoms in each frequency band is counted and is used as a feature F_n. This gives precise estimate of the contribution of TF functions to approximate particular frequency components of any given signal.

$$F_n = Numberofatoms(B_n) \quad (5)$$

where B_n is the frequency band and $n = 1, 2, \ldots, N(= 12)$.

For a given signal, atoms are obtained using MPTK [10] in MATLAB environment with number of iterations, M = 2000 and scale factor, $s_m = \{2^1, 2^2, \ldots, 2^{13}\}$ in eq 2. Five pathological groups are considered here: adductor spasmodic dysphonia, keratosis, paralysis, vocal nodules and vocal polyps. The number of wave files in each group is tabulated in table II and are taken from MEEI [11] database. The sampling frequency of few files was at 50kHz and it is down sampled to 25kHz. Such that all the wave files are at 25kHz. All the files are amplitude normalised and no other processing steps are applied before applying matching pursuit algorithm.

B. Feature extraction by PRAAT

Jitter and shimmer are the most widely used perturbation measures for pathological voices. Here, jitter, shimmer and its variants are derived using pitch marks obtained by PRAAT. We use the default setting of PRAAT for pitch mark extraction and MATLAB environment to compute jitter, shimmer, and its derivatives by the extracted pitch marks. The default settings like, the Period Floor (PF) and Period Ceiling (PC) parameters are set at 50Hz and 10kHz respectively, Maximum Amplitude Factor (MAP) and Maximum Period Factors (MPF) are set to 1.6 and 1.3 respectively. While pitch mark extraction, PF and PC parameters are used and for feature extraction, all the four (PF, PC, MAF and MPF) threshold parameters are used. The following explains the equations used for feature extraction.

TABLE I: Frequency Bands

Band	Frequency
Band 1	<1kHz
Band 2	1 to 2 kHz
Band 3	2 to 3 kHz
Band 4	3 to 4 kHz
Band 5	4 to 5 kHz
Band 6	5 to 6 kHz
Band 7	6 to 7 kHz
Band 8	7 to 8 kHz
Band 9	8 to 9 kHz
Band 10	9 to 10 kHz
Band 11	10 to 11 kHz
Band 12	11 to 12.5 kHz

1) Absolute Jitter : It is the evaluation of the period to period variability of the pitch period within the analysed voice sample and is computed by

$$absolutejitter(\text{seconds}) = \frac{1}{N-1} \sum_{i=1}^{N-1} \left| To^{(i)} - To^{(i+1)} \right| \quad (6)$$

where $To^{(i)}$ is the duration of the ith interval and N is the number of intervals.

2) Jitter: It is a measure of period to period variability of the pitch within the analysed voice sample and is computed by

$$jitter = \frac{\frac{1}{N-1} \sum_{i=1}^{N-1} \left| To^{(i)} - To^{(i+1)} \right|}{To} \quad (7)$$

where N is the number of extracted pitch marks, $To = \frac{1}{N} \sum_{i=1}^{N} To^{(i)}$ and $To^{(i)}$ is the duration of the i-th pitch period.

Pitch period perturbation (PPQ), relative average perturbation (RAP), and smoothed pitch period perturbation (sPPQ) are computed similar to jitter but with 3, 5 and 55 pitch cycles respectively. The description and equations are available in Multi-Dimensional Voice Program (MDVP) manual [11] and the same is used here.

3) Shimmer in db : It is the evaluation of the period to period variability of the peak to peak amplitude in dB within the analysed voice sample and is computed by

$$absoluteshimmer(\text{db}) = \frac{1}{N-1}\sum_{i=1}^{N}\left|20log(A^{i+1}/A^{i})\right| \tag{8}$$

where $A^{(i)}, i = 1, 2, \ldots, N$ - extracted peak-to-peak amplitude N- number of extracted impulses.

4) Shimmer: It is a measure of period to period variability of the peak to peak amplitude within the analysed voice sample and is computed by

$$shimmer = \frac{\frac{1}{N-1}\sum_{i=1}^{N-1}\left|A^{(i)} - A^{(i+1)}\right|}{\frac{1}{N}\sum_{i=1}^{N}A^{(i)}} \tag{9}$$

where $A^{(i)}, i = 1, 2, \ldots, N$ are the extracted peak-to-peak amplitude and N is number of extracted pitch marks.

Amplitude perturbation quotient (APQ) and smoothed amplitude perturbation quotient (sAPQ) are computed similar to shimmer but with 5 and 55 peak-to-peak cycles. The description and equations are available in MDVP manual [11] and the same is used here.

5) Coefficient of fundamental frequency variation in time (vTo) : It is a relative standard deviation of the fundamental frequency in time. It reflects the variation of To within the analysed voice sample and is computed by [11],

$$vTo = \frac{\sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(\frac{1}{N}\sum_{j=1}^{N}To^{(j)} - To^{(i)}\right)^{2}}}{To} \tag{10}$$

where N is the number of extracted pitch marks, $To = \frac{1}{N}\sum_{i=1}^{N}To^{(i)}$ and $To^{(i)}$ is the duration of the i-th pitch period.

6) Coefficient of amplitude variation (vAm) : It is a relative standard deviation of the peak to peak amplitude. It reflects the peak to peak amplitude variations within the analysed voice sample and is computed by [11],

$$vAm = \frac{\sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(\frac{1}{N}\sum_{j=1}^{N}A^{(j)} - A^{(i)}\right)^{2}}}{\frac{1}{N}\sum_{i=1}^{N}A^{(i)}} \tag{11}$$

where $A^{(i)}, i = 1, 2, \ldots, N$ are the extracted peak-to-peak amplitude and N is number of extracted pitch marks.

7) The standard deviation (STD, σ) of fundamental frequency in time ($To^{(i)}$) is computed by:

$$\sigma = \sqrt{\frac{1}{N}\sum_{i=1}^{N-1}\left(To - To^{(i)}\right)^{2}} \tag{12}$$

where N is the number of extracted pitch marks, $To = \frac{1}{N}\sum_{i=1}^{N}To^{(i)}$ and $To^{(i)}$ is the duration of the i-th pitch period.

IV. Classification and Results

Feature classification is done using random forest classifier (available in MATLAB 2011a), with uniform prior probability for each class and Gini's diversity index split criterion. The model is cross-validated using leave-one-out method. In this method, one sample is excluded for training and the excluded sample is used for testing. This is repeated until all the files are tested against the rest of files. True positive (TP) is defined as the ratio of number of files classified correctly as pathological type say P1 to the total number of P1 type pathological files. Similar to TP, true negative (TN) is defined as the ratio of number of files NOT classified as pathological type P1 and total number of pathological files other than P1 type. Overall score is the ratio of number of files classified as pathological type P1 and number of files classified NOT as pathological type P1 to the total number of files.

The MEEI database [11] is used to test the features discussed in section III using Random Forest classifier with leave-one-out cross validation method. The number of files available in each pathological subgroup types in MEEI database, which are considered here are tabulated in table II. Classification scores are shown in table III, IV and V for adductor, keratosis and vocal nodules respectively.

It can be seen in table III that for adductor, MP shows highest classification score in Band 12 than in the other bands which signifies that the number of atoms to approximate signal's frequency components in 11kHz to 12.5kHz has more classification power than the rest of the bands. In addition to Band 12, Band 11 and Band 4 shows classification accuracy comparable to Band 12. In case of PRAAT, absolute jitter, jitter and coefficient of amplitude variation shows good classification score and is comparable with the scores obtained in Band 4, Band 11 and Band 12. In case of keratosis in table IV, Band 3 and Band 12 of MP are showing highest classification score compared to other bands. For PRAAT, coefficient of amplitude variation and shimmer in db are showing reasonably good scores compared to other features and the performance of MP features are better than PRAAT. Similar to adductor and keratosis, in case of vocal nodules in table V, Band 7 of MP is showing the highest classification score when compared to other bands. For PRAAT, standard deviation of the fundamental frequency in time and coefficient of fundamental frequency variation in time is showing highest classification score compared to all the other features.

Vocal nodules classification score (for Band 7 in MP and for STD in PRAAT) is highest of all the three pathological types considered here. Although STD can better classify vocal

TABLE II: Pathological type and number of files

Pathological type	Number of files
Adductor	19
Keratosis	27
Paralysis	67
Vocal Nodules	19
Vocal Polpys	20

nodules than Band 7, STD depends on the efficiency of the pitch mark extraction algorithm. Greater classification score definitely doesn't imply that the pitch mark algorithm is better since we are dealing with the pathological voices which are non-stationary in nature. Also by manual observation, it can be known that PRAAT performance in pitch mark extraction can be believed only for normal voices. On contrary, Band 7 classification score is more reliable since this feature is obtained by time-frequency localisation method i.e. Matching pursuit. On similar lines, for adductor and keratosis also, scores of MP features can be considered as more stable than the PRAAT features.

TABLE III: Classification of adductor

Matching Pursuit (MP)				PRAAT			
Bands	**TP**[%]	**TN**[%]	**Overall**[%]	**Features**	**TP**[%]	**TN**[%]	**Overall**[%]
Band 1	40	67.71	63.94	**abs_jitter**	40	75.93	71.24
Band 2	20	66.92	60.54	**jitter**	45	75.18	71.24
Band 3	35	67.71	63.26	**RAP**	30	72.18	66.67
Band 4	45	74.01	70.06	**PPQ**	50	71.42	68.62
Band 5	15	53.54	48.29	**sPPQ**	20	70.67	64.05
Band 6	10	63.77	56.46	**db_shimm**	30	72.18	66.67
Band 7	20	59.05	53.74	**shimmer**	31	64.66	59.47
Band 8	15	66.14	59.18	**APQ**	32	73.68	66.01
Band 9	10	51.18	45.57	**sAPQ**	33	72.93	67.97
Band 10	55	70.86	68.70	**T0_var**	34	67.66	63.39
Band 11	25	77.16	70.06	**amp_var**	35	75.93	71.24
Band 12	40	76.37	71.42	**STD**	36	70.67	66.67

V. Conclusion

Matching pursuit algorithm is used to classify pathological sub-groups consisting of adductor spasmodic dysphonia, keratosis, paralysis, vocal nodules and vocal fold polyps. Frequency band of the atoms obtained by MP decomposition is used as a feature. Classification score shows encouraging results due to good time-frequency localisation of MP algorithm. This approach has shown highest classification score in case of vocal nodules than the rest of the pathological types. Using

TABLE IV: Classification of keratosis

Matching Pursuit (MP)				PRAAT			
Bands	**TP**[%]	**TN**[%]	**Overall**[%]	**Features**	**TP**[%]	**TN**[%]	**Overall**[%]
Band 1	22.23	72.8	63.81	**abs_jitter**	29.62	65.07	58.82
Band 2	25.92	71.2	63.15	**jitter**	18.51	67.46	58.82
Band 3	44.45	71.2	66.44	**RAP**	18.51	65.07	56.86
Band 4	18.51	60.8	53.28	**PPQ**	22.23	63.49	56.20
Band 5	7.40	65.6	55.26	**sPPQ**	22.23	60.31	53.59
Band 6	29.62	60	54.60	**db_shimm**	25.92	70.63	62.74
Band 7	33.34	69.6	63.15	**shimmer**	33.34	61.12	56.20
Band 8	29.62	64.8	58.55	**APQ**	33.34	57.14	52.94
Band 9	33.34	64.8	59.21	**sAPQ**	18.51	47.61	42.48
Band 10	37.03	56.8	53.28	**T0_var**	22.23	59.52	52.94
Band 11	59.25	48.8	50.65	**amp_var**	29.62	72.23	64.70
Band 12	29.62	76	67.76	**STD**	25.92	66.67	59.47

TABLE V: Classification of vocal nodules

Matching Pursuit (MP)				PRAAT			
Bands	**TP**[%]	**TN**[%]	**Overall**[%]	**Features**	**TP**[%]	**TN**[%]	**Overall**[%]
Band 1	15.78	72.86	65.54	**abs_jitter**	21.05	78.35	71.24
Band 2	26.31	75.96	69.59	**jitter**	57.89	76.86	74.50
Band 3	52.63	76.74	73.64	**RAP**	21.05	76.86	69.93
Band 4	26.31	67.44	62.16	**PPQ**	36.84	78.35	73..20
Band 5	21.05	64.34	58.78	**sPPQ**	31.57	80.59	74.50
Band 6	36.84	65.89	62.16	**db_shimm**	31.57	76.86	71.24
Band 7	63.15	80.62	78.37	**shimmer**	36.84	77.61	72.54
Band 8	57.89	62.79	62.16	**APQ**	21.05	73.88	67.32
Band 9	57.89	61.24	60.81	**sAPQ**	21.05	72.38	66.01
Band 10	64.03	70.54	68.24	**T0_var**	47.36	82.08	77.77
Band 11	31.57	58.13	54.72	**amp_var**	15.78	71.64	64
Band 12	52.63	61.24	60.13	**STD**	57.89	88.05	84

MP is advantageous over PRAAT because it doesn't rely on the pitch information. Any advanced decomposition method with better time frequency locational may perform better than MP. There is much scope to explore in this pathological subgroup classification, which is more challenging than general normal and pathological voice classification.

References

[1] Markaki, M., Stylianou, Y: "Voice Pathology Detection and Discrimination Based on Modulation Spectral Features," IEEE Transactions on Audio, Speech, and Language Processing, Vol 19, pp. 1938 - 1948 (2011).

[2] A. Tsanas, M.A. Little, P.E. McSharry, J. Spielman, L.O. Ramig: "Novel speech signal processing algorithms for high-accuracy classification of Parkinson's disease," IEEE Transactions on Biomedical Engineering, Vol. 59, pp. 1264-1271 (2012).

[3] Umapathy, K., Krishnan, S., Parsa, V., Jamieson, D.G: "Discrimination of pathological voices using a time-frequency approach," IEEE Transactions on Biomedical Engineering, Vol. 52, pp. 421-430 (2005).

[4] Ce Peng,Wenxi Chen, Zhu Xin, Baikun Wan, Daming Wei: "Pathological Voice Classification Based on a Single Vowel's Acoustic Features," 7th IEEE International Conference on Computer and Information Technology, pp. 1106-1110 (2007).

[5] Markaki M, Stylianou Y: "Using modulation spectra for voice pathology detection and classification," IEEE International Conference on Engineering in Medicine and Biology Society, pp. 2514 - 2517 (2009).

[6] Hosseini, P.T., Almasganj, F., Emami, T., Behroozmand, R., Gharibzade, S., Torabinezhad, F: "Local Discriminant Wavelet Packet Basis for Voice Pathology Classification," IEEE 2nd International Conference on Bioinformatics and Biomedical Engineering, pp. 2052 - 2055 (2008).

[7] Boersma, Paul and Weenink, David. Praat: doing phonetics by computer [Computer program]. Version 5.4.08, retrieved 24 March 2015 from http://www.praat.org/ (2015).

[8] A. Tsanas, M. Zaartu, M.A. Little, C. Fox, L.O. Ramig, G.D. Clifford: "Robust fundamental frequency estimation in sustained vowels: detailed algorithmic comparisons and information fusion with adaptive Kalman filtering," Journal of the Acoustical Society of America, 135, 2885 (2014)

[9] Khalid Daoudi, Blaise Bertrac, "On classification between normal and pathological voices using the MEEI-KayPENTAX database: Issues and consequences" INTERSPEECH-2014, Singapore (2014).

[10] Krstulovic, S. and Gribonval, R. "MPTK: Matching Pursuit made Tractable," in Proc. ICASSP (3):496-499 (2006).

[11] M. Eye and E. Infirmary, "Voice Disorders Database," Version 1.03. Lincoln Park, NJ: Kay Elemetrics Corporation (1994).