INTELLIGENT VOICE QUALITY ASSESSMENT POST-TREATMENT USING GENETIC PROGRAMMING

Walaa Sheta[1], Tim Ritchings[2] & Carl Berry[2]

[1]Mubarak City for Scientific Research,
Burg El-Arab, Egypt
wsheta@mcit.gov.eg

[2]School of Computing, Science and Engineering,
University of Salford, UK

Abstract: **Objective techniques for assessing and classifying voice quality for patients recovering from treatment for cancer of the larynx have largely focussed on they use of Artificial Neural Networks (ANN). The results of a preliminary study are reported, where a Genetic Programming (GP) has been trained to classify recovered (normal) and abnormal voices in acoustic data, and produced much more accurate results than an ANN. In addition, the GP is able to provide impact factors for the various voice parameters, and suggests that only 6 of the 22 short-term and long-term parameters used in the current ANN studies are contributing significantly to the classifications.**

Keywords: Voice quality, classification, Artificial Neural Network, Genetic Algorithms, acoustic signals.

I. INTRODUCTION

The use of intelligent computer-based techniques to support decision making in clinical applications, have been investigated over the years for a wide variety of clinical data. Although Genetic Programming (GP) has not been used extensively for medical applications to date, the early results for cancer diagnosis [1,2] were found to be better than with an Artificial Neural Network (ANN). In another study [3], a grammar-based GP variant was used for knowledge extraction from medical databases, where the rules for the diagnosis were derived from an algorithm that uncovers relationships among data attributes. The outcomes of different types of classifiers, including ANNs and genetic programs have also been reported [4].

This study is part of a larger project which is concerned with developing objective techniques for voice quality assessment in patients recovering from cancer of the larynx. The earlier investigations have concentrated on the use of Artificial Neural Networks (ANN) to firstly distinguish recovered (normal) and abnormal voices [5] on the basis of a collection of short-term and long-term parameters derived from the patient's voice signals, and more recently, classify voices on the 7-point scale for voice quality used by Speech and Language Therapists (SALT) in the UK [6]. Similar classifications have now been obtained for both electrical impedance (EGG) and acoustic data, with the best results for voices in the extremes categories on this scale (normal and abnormal), while those for the mid-categories have been poor [7].

A preliminary assessment of use of Genetic Programming to classify normal (recovered) and abnormal acoustic signals is described here. The resulting classifications are compared with those obtained from an ANN for the same signal parameters and training regimes..

II. TREATMENT OF VOICE SIGNALS

A. Collection of Voice Signals

The patient's voice data was collected by the Christie Hospital and the South Manchester hospital using an electrolaryngograph PCLX system [8]. The equipment simultaneously records the electrical impedance signal via pads placed at specific positions on the patient's neck at the same time as the acoustic voice signal using a microphone. In these studies, the patient was attempts to steadily phonate the /i/ sound. Although two datasets are collected, only the acoustic data have been used to date in this study. In the work only the male voices were used as the number of female voices in the dataset was too small to give an accurate assessment, a feature of the dataset is that most cancer of the larynx patients are male.

Voice quality was subjectively classified by a SALT for each patient using their standard 7-point classification scale ranging from Lx0-Lx6, with Lx0 being a near normal (recovered) voice while Lx6 represents an abnormal, very poor quality voice. The approach taken to reach a classification is very subjective and depends to a large extent on the experience of the SALT In this study the Lx0 and Lx1 voices were combined and considered as the normals, while Lx5 and Lx6 voices were combined to give the abnormals, The number of patients in these two categories is shown in Table 1.

Models and analysis of vocal emissions for biomedical applications. 4th international workshop. October 29-31, 2005 – Firenze, Italy.
Edited by C. Manfredi. ISBN 88-8453-320-1 (online) © 2005 Firenze University Press

Normal (Lx0,Lx1)	Abnormal (Lx5,Lx6)
58	36

Table 1. Patient numbers used in this study

B. Signal Pre-processing

In order to be able to extract the short and long term parameters used in the classification process, a number of pre-processing stages were applied to the voice signals. Initially the signals were stationarised to remove drift, split into 50 ms frames (Hanning windows overlapping by 25 ms) and then converted to the autocorrelation form of the signal to remove some of the noise components. Once these processes were complete, the frames were examined to check if they contained silence or sound. This involved comparing the frames with a sample of silence frame recorded under the same conditions, and used zero point crossing and short term amplitudes as checks. Once the silence frames have been removed, the remaining frames were separated into voiced and unvoiced frames; voiced frames containing vocal phonation while unvoiced containing no recognisable speech. This was achieved using the cepstrum based approach as described in [9]. The Fundamental Harmonic Normalisation (FHN) as described in [10] was then calculated from Power Spectrum Density (PSD) and then this structure was modelled by fitting a Gaussian Mixture Model (GMM) in order to reduce the number of parameters needed to describe the signal.

C. Parameter Extraction

A total of 22 short-term and long-term parameters are extracted for use with classification, as detailed in [5,6]. The short term parameters consist of 15 parameters relating to the mean, standard deviation and peak of the gaussians used to describe the fundamental frequency and first four harmonics in the frame (if they can be detected)) ($M_{0..4}$, $SD_{0..4}$, $P_{0..4}$); the value of the fundamental frequency in each frame (F_0), the noise threshold value (N_0), the FHN Noise Energy (FHNNE) and the Residual Harmonic Energy (RHE). The 3 long-term parameters were extracted from the speaker's whole voiced speech. These included the mean fundamental frequency across all frames (MF_0), a measure of jitter of the fundamental frequency between frames (J_0) and the ratio of voiced to unvoiced frames (VS).

*D. The GP classification techniqu*e

Linear Genetic Programming was used to classify the normal and abnormal voices. An experiment with 7 runs was performed using this technique, the runs only differing in their choice of a random seed. The common parameter settings used in the experiment are given in Table 2.

Parameter	Value
Population size	512
Max no of tournaments	150000
Mutation frequency	30
Crossover frequency	30
Max program size	256
Instruction set	+ - * / sin() log()

Table 2. Parameter settings for the GP

All 22 short-term and long-term parameters were extracted from the voice signals and used for classification. The dataset was split into a training (65%) and test (35%) set, which equated to 38/20 for the normals and 23/13 for the abnormals.

*E The ANN classification techniqu*e

The same 22 parameters were used for the GP classification and the same 65/35% spilt for the training and test data sets. In this case, the parameters were input to 3 layer feed-forward ANN with a sigmoidal activation function in the hidden layer. Two different training algorithms were used; gradient descent with momentum backpropagation (TRAINGDM), and resilient backpropagation (TRAINRP). The results were not found to be dependent on the actual number of hidden nodes.

3. RESULTS AND DISCUSSION

A. Classifications using the full parameter set.

The results obtained when the 22 short-term and long-term parameters were used by the GP and the ANN are given in Table 3.

	Normal (Lx0,Lx1)	Abnormal (Lx5,Lx6)
GP	99.6±2.4%	97.2±2.9%
ANN	90.2±2.1%	87.5±3.9%

Table 3. Classification accuracies using the GP and ANN

The classifications for the ANN were slightly lower that those obtained using the "leave one out" cross validation strategy which is generally regarded as one of the most accurate methods and by leaving out a single patient's voice sample we can ensure to avoid inter versus intra speaker effects [6]. However, the GP was clearly found to give the more accurate classifications.

B. *Classifications using the impact parameter set.*

One of the advantages of using the GP is that it provides the impact factor of each parameter on the classification. Table 4 shows how each parameter contributed in the generated program. The Table shows the frequency percentage of the best thirty generated programs containing the referenced input; the average effect of removing all instances of that input, and the maximum impact of that input. In these cases, the greater the value, the more impact removal of that input had.

Parameter	Frequency	Average Impact	Maximum impact
VS	1.00	13.69	17.78
MF_0	0.67	7.41	10.50
J_0	0.77	4.45	9.78
N_0	0.23	1.13	2.20
M_0	0.17	0.32	0.45
P_0	0.27	0.27	0.27
RHE	0.17	0	0
SD0	0.10	0	0
P_1	0.10	0	0
SD_3	0.10	0	0
P_3	0.10	0	0
M_1	0.07	0	0
SD_1	0.07	0	0
P_4	0.07	0	0
FHNNE	0.07	0	0
SD_2	0.03	0	0
P_2	0.03	0	0
M_4	0.03	0	0
M_2	0	0	0
M_3	0	0	0
SD_4	0	0	0
F_0	0	0	0

Table 4. Contribution of each parameter

It may be seen from the Table that only 6 of the 22 parameters were found to have a significant impact on the classification. These parameters were N_0, M_0, P_0, MF_0, J_0, and VS.

Both the GP and the ANN were re-trained and tested using just these 6 parameters, and the results are shown in Table 5.

	Normal (Lx0,Lx1)	Abnormal (Lx5,Lx6)
GP	99.2±3.1%	96.4±3.7%
ANN	88.6±2.7%	81.5±4.2%

Table 5. Classification accuracies using GP and ANN

V. CONCLUSIONS.

A preliminary study has been made involving the use of GP to classify recovered (normal) voices and abnormal voices in acoustic signals taken from patients recovering from cancer of the larynx. Initially, a collection of 22 short-term and long-term parameters were extracted from the signal and used as input to the GP, and also an ANN. The GP provided much more accurate classifications than the ANN.

Examination of the impact factors for the voice parameters suggests that there are only 6 significant factors. The results obtained from both the GP and the ANN using just these parameters were only slightly poorer than for the full parameter set, again with the GP providing the more accurate classifications.

One of the advantages of the ANN is the ability to produce multiple outputs, enabling classifications to be made corresponding to the 7-point scale for voice quality used by SALTS. Work is now taking place to extend the GP approach to multiple classifications.

REFERENCES

[1] Gray, H., Maxwell, R., Martinez-Perez, I., Arus, C. and Cerdan, S. "Genetic programming for classification of brain tumours from nuclear magnetic resonance biopsy spectra ,in Genetic Programming", 1st *International Conference on Genetic Programming, GP-96,* Stanford, US, 1996.

[2] Nordin, P., Keller, R., and Francone, F. "Genetic Programming: an introduction on the automatic evolution of computer programs and its applications." Wolfgang Banzhaf Publishers, Inc. 2002.

[3] Ngan, P., M. Wong, M., Leung, K. and Cheng, J. "Using grammar based genetic programming for data mining of medical knowledge, in Genetic Programming" 3rd *International Conference on Genetic Programming, GP-98,* Madison, US, 1998.

[4] Somorjai, R., Nikulin, A., Pizzi, N., Jackson, D., Scarth, G., Dolenko, B., Gordon, H., Russell, P., Lean, C., Delbridge, L., Mountford, C. and Smith, I. "Computerized consensus diagnosis - A classification strategy for the robust analysis of MR spectra1. Application to H-1 spectra of thyroid neoplasma", *Magn. Reson. Med.* **33**, *257–263,* 1995.

[5] Ritchings RT, McGillion M, Moore CJ "Pathological voice quality assessment using artificial neural networks." *Medical Engineering and Physics* 24 (2002), pp561-564, PII S1350-4533(02)00064-4.

[6] Godino-Llorente, JI, Ritchings, RT, Berry C "The Effects of Inter and Intra Speaker Variability on Pathological Voice Quality Assessment" *3rd International Workshop on Models and Analysis of Vocal Emissions for Biomedical Applications*, Firenze, Italy 2003

[7] Berry, C and Ritchings, RT. "A comparative study of intelligent voice quality assessment using impedance and acoustic signals.*" 4th International Workshop on Models and Analysis of Vocal Emissions for Biomedical Applications,* Firenze, Italy 2005.

[8] Fourcin, A.J., Abberton E., Miller, D, Howell D. Laryngograph : "Speech pattern element tools for therapy, training and assessment." *European Journal of Disorders of Communication* 30(2), 1996, pp.101-115

[9] Rabiner, L. and Juang, B.H. Fundamentals of speech recognition. New Jersey Prentice Hall, 1993.

[10] Moore C.J., Manickam K., Slavin N. "Voicing recovery in males following radiotherapy for larynx cancer.*" 4th International Workshop on Models and Analysis of Vocal Emissions for Biomedical Applications,* Firenze, Italy 2005.

.