<table>
<thead>
<tr>
<th><strong>Title</strong></th>
<th>Acoustical analysis of voices produced by Cantonese patients of unilateral vocal fold paralysis: acoustical analysis of voices by Cantonese UVFP</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Author(s)</strong></td>
<td>Yan, N; Wang, L; Ng, ML</td>
</tr>
<tr>
<td><strong>Citation</strong></td>
<td>The 2013 IEEE International Conference on Signal Processing, Communications and Computing (ICSPCC 2013), Kunming, Yunnan, China, 5-8 August 2013. In Conference Proceedings, 2013</td>
</tr>
<tr>
<td><strong>Issued Date</strong></td>
<td>2013</td>
</tr>
<tr>
<td><strong>URL</strong></td>
<td><a href="http://hdl.handle.net/10722/205849">http://hdl.handle.net/10722/205849</a></td>
</tr>
<tr>
<td><strong>Rights</strong></td>
<td>IEEE International Conference on Signal Processing, Communication and Computing (ICSPCC). Copyright © IEEE.; ©2013 IEEE. Personal use of this material is permitted. However, permission to reprint/republish this material for advertising or promotional purposes or for creating new collective works for resale or redistribution to servers or lists, or to reuse any copyrighted component of this work in other works must be obtained from the IEEE.; This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License.</td>
</tr>
</tbody>
</table>
Acoustical analysis of voices produced by Cantonese patients of unilateral vocal fold paralysis

Nan Yan, Lan Wang
Laboratory of Ambient Intelligence and Multimodal Systems
Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences/The Chinese University of Hong Kong
Shenzhen, China
nan.yan@siat.ac.cn lan.wang@siat.ac.cn

Manwa L. Ng
Speech Science Laboratory, Division of Speech and Hearing Sciences
University of Hong Kong
Hong Kong, China
manwa@hku.hk

Abstract—Injection laryngoplasty using hyaluronic acid for patients of unilateral vocal fold paralysis (UVFP) is becoming more prevalent in Hong Kong. The present investigation explored the efficacy of the procedure by examining various acoustical measures. Both traditional perturbation analyses and nonlinear dynamical measures were used to assess the change in patients’ voice quality between before, one month and three months after injection laryngoplasty. Jitter, shimmer, noise-to-harmonic ratio (NHR), and nonlinear recurrence period density entropy (RPDE), detrended fluctuation analysis (DFA) and correlation dimension (D2) were obtained from the sustained vowel /i/ produced by ten native Cantonese UVFP patients. Results showed that RPDE and D2 were sensitive to improvement after injection laryngoplasty. The receiver operating characteristic area under curve analysis (ROC-AUC) was also used to demonstrate the diagnostic performance of these objective measures. High specificity of these four acoustic measures in distinguishing the voice quality before and after surgery was found, indicating that nonlinear dynamical analysis could be a supplemental tool to traditional acoustic analysis, and if combined, an effective monitoring tool for UVFP treatment.

Index Terms—unilateral vocal fold paralysis, acoustical analysis, recurrence period density entropy, correlation dimension, detrended fluctuation analysis.

I. INTRODUCTION

Unilateral vocal fold paralysis (UVFP) is a vocal pathology caused by lesion of the recurrent laryngeal nerve on one side. UVFP patients usually exhibit signs of hoarseness, breathiness, inadequate coughing, and feeling of fatigue and sometimes pain in the neck [1]. The perceptually poor quality of the UVFP voices is mainly due to the distorted glottic anatomy causing a change in biomechanical and aerodynamical properties of glottal vibration. Although some speech therapy intervention appears to help with the condition [2], surgery such as injection laryngoplasty, medialization thyroplasty, arytenoid adduction, and laryngeal reinnervation is still the primary treatment approach for UVFP. Among them, injection laryngoplasty using different types of filling materials is becoming more common for UVFP including Hong Kong [3, 4], thanks to its simplicity, reversibility and predictability of outcomes [5-7]. The immobility of one vocal fold results in a glottal gap upon vocal adduction. To compensate the glottic gap, in injection laryngoplasty, implant material is injected to the medial edge of the paralyzed fold, in an attempt to close the gap between the two vocal folds. The procedure is done with local anesthesia and the patient is able to phonate during the procedure [7].

To quantitatively examine treatment efficacy, acoustical parameters have been used to assess the change of voice quality [8]. Perturbation analyses have often been used in clinical practice such as analyzing alaryngeal or dysphonic voices. The traditional acoustical parameters including jitter, shimmer and noise-to-harmonic ratios (NHR) are mostly based on waveform-based cycle analysis, and linear digital signal processing. However, the algorithms heavily rely on the assumption of signal linearity, rendering practical limitations when using such analyses.

According to Titze et al. [9], voice signals can be classified into three categories: (1) nearly-periodic signals, (2) signals with strong subharmonics or modulations, and (3) aperiodic signals. They argued that traditional acoustical methods such as jitter and shimmer analysis was suitable only for Type 1 speech signals. Compared to classical analysis based on linear framework, acoustic analyses using nonlinear dynamics techniques are simpler and have already been successfully applied to analyses of other pathological voices [10-12]. Previous studies [11, 13] examined performance and reliability of nonlinear dynamical analysis using a variety of vocal pathologies compared with classical perturbation methods. Yan et al. found that the nonlinear dynamical analysis is more reliable than traditional perturbation measures when using a different lengths of aperiodic voice signals [13]. Zhang et al. studied various vocal pathologies such as vocal polyps and unilateral laryngeal paralysis [11, 14]. They noted the
sensitivity and specificity associated with the method in distinguishing patients from normal subjects. However, few studies have used such acoustical analytic method to evaluate treatment efficacy of surgical intervention for UVFP. Only one study reported examination of the performance of nonlinear measures in accessing the effectiveness of medialization thyroplasty surgery in treating vocal paralysis [10]. They observed that nonlinear measures can be useful as a treatment effectiveness monitoring tool for UVFP. Yet, few studied injection laryngoplasty, and most of them only reported surgical outcomes of injection laryngoplasty from English speaking patients. Although injection laryngoplasty is increasingly common for patients with glottic incompetence in Hong Kong, no study has examined its efficacy regarding vocal outcome in the Cantonese population. It is not known if patients could equally benefit from the procedure.

The present study served as an extension to Litters et al.’s study [10] by investigating the efficacy of injection laryngoplasty based on a tone language (Cantonese). Several nonlinear dynamical parameters were obtained from UVFP voice signals produced by patients who were native Cantonese speakers living in Hong Kong both before and after receiving the treatment of injection laryngoplasty using hyaluronic acid. The parameters included correlation dimension (D2), recurrence period density entropy (RPDE), and detrended fluctuation analysis (DFA). Traditional perturbation measures were also obtained from these voice signals and used to compare with nonlinear dynamic measures.

II. METHODS

A. Participants

Ten (six male and four female) adult patients who were diagnosed with UVFP participated in the study. They underwent injection laryngoplasty using hyaluronic acid (Restylane) carried out by an experienced otorhinolaryngologist. The inclusion criteria of patients were: (1) they were native speakers of Cantonese; (2) all patients were between 40-70 years of age and have no other history of speech and/or hearing problem, except that associated with the UVFP; (3) they were referred by speech therapists; and (4) the patients had completed a course of speech therapy and showed no improvement in voice. Only patients with idiopathic/iatrogenic UVFP who showed spontaneous recovery were recruited. All patients were willing to participate in the study.

B. Injection laryngoplasty procedure

Patients recruited were arranged for injection laryngoplasty that was carried out in the Ear, Nose, and Throat Clinic of the Queen Mary Hospital. On the day of procedure, the patient was asked to fast for six hours and then assessed by using video stroboscopy and flexible laryngoscopy to confirm the diagnosis. Local anesthesia was administered to the larynx and nearby skin tissue. Hyaluronic acid (Restylane) was injected percutaneously by the otorhinolaryngologist to medialize the paralyzed fold and to close the interglottal gap which was simultaneously examined through the flexible laryngoscope held by another assistant otorhinolaryngologist who was also experienced in the procedure. The patient was discharged home one hour after the procedure and feeding could be resumed.

C. Speech tasks and recording procedure

The speech task included vowel prolongation. During the experiment, the participants were instructed to sustain the vowel /i/ at high level tone three times for as long as they could. With no attempt to control for loudness, the UVFP patients were instructed to produce the speech samples at a comfortable level of loudness.

All measurements were obtained three times: once before injection was carried out, then one month, and three months post injection. A brief instruction of the recording procedure was provided before the recording took place. To avoid recording of extraneous noise, all recordings took place in a sound-treated booth. During the recording, the microphone was placed approximately 10 cm from the speakers’ mouth. Acoustic signals were obtained using a high-quality microphone (SM58, Shure, USA) via a professional grade pre-amplification system (MobilePre USB, M-Audio, USA). The recorded acoustic signals were digitized using Praat at a sampling rate of 20 kHz and quantization rate of 16 bits/sample.

To avoid the effect from phonation initiation and termination, only the medial 80% of the sustained vowel segment was used for analysis. For non-linear dynamical analysis, a half-second segment was selected from this vocalic portion.

D. Perturbation analysis

Jitter, shimmer and NHR values were obtained from the voice samples. Jitter is a measure of short-term (cycle-to-cycle) variation in the fundamental frequency of a signal, whereas shimmer measures the amplitude variation of a signal. The NHR was calculated as the average inverse harmonics-to-noise ratio (HNR) over each cycle. All perturbation measures were calculated using the Praat software system [15].

E. Nonlinear dynamical analysis

The dynamics of each voice segment was reconstructed in a phase space, which was then used to calculate correlation dimension (D2) and recurrence period density entropy (RPDE). D2 specifies the degree of freedom needed to describe a system. When a more complex system has a higher D2, more degrees of freedom are needed to describe its dynamical state [14]. RPDE estimates the relative uncertainty in the expected recurrence periods of the embedded signal. That is, the time intervals between successive close returns to the same point in the space spanned by the lagged vectors. It can be shown that this generalizes the notion of cycle length for cyclic signals, and period for exactly periodic signals [10]. In the present study, an m-dimensional delay coordinate phase space \( X_\tau = \{ x(t), x(t-\tau), ..., x(t-(m-1)\tau) \} \) was reconstructed using the time delay technique, where \( m \) is the embedding dimension and \( \tau \) the time delay. \( m \) was determined using the False Nearest Neighbors method [16] and the proper time delay \( \tau \) was estimated using the C-C methods [17]. Correlation integral \( C(r) \) measures the number of distances between points in the reconstructed phase space that
are smaller than the radius $r$. $D_2$ and RPDE were manually estimated in the scaling region of the radius $r$ with the embedding dimension $m$.

DFA characterizes the extent of turbulent noise in the voice signal, quantifying the stochastic self-similarity of the noise caused by turbulent airflow in the vocal tract. Breathiness or other similar dysphonia caused by vocal pathology such as incomplete vocal fold closure can increase the DFA value [10]. Reliability of all these nonlinear measures has been demonstrated in previous studies [10, 18].

**F. Statistical analysis**

Since the nonlinear parameters have non-Gaussian population, nonparametric Wilcoxon signed-rank test was used to assess perturbation and nonlinear measures were significantly different among the voice types (independent variable).

**III. RESULTS**

The typical waveform and the reconstructed phase spaces of pre-op UVFP and post-op UVFP voices are illustrated in Fig. 1. Pre-op and post-op voices are associated with an aperiodic waveform (Figs. 1a, 1c and 1e). As shown in Figs. 1b, 1d and 1f, the reconstructed phase spaces of UVFP voices consistently depict an irregular structure.

Table 1 shows the summary statistics for the perturbation and nonlinear dynamical measures. Both median and mean of all measures consistently indicated differences between before and after surgery. The corresponding statistical results reflected detailed inter-category differences. To quantify the effect size, receiver operating characteristic area under curve (ROC-AUC), which indicated the amount of overlap between categories, was calculated.

Results of Wilcoxon signed-rank test indicated significant changes in RPDE ($Z = -1.988, p = 0.047$), $D_2$ ($Z = -2.497, p = 0.013$) and shimmer values ($Z = -2.100, p = 0.036$) one month after surgery. At three months after surgery, significant differences were found in RPDE ($Z = -2.497, p = 0.013$), $D_2$ ($Z = -2.497, p = 0.013$) and shimmer ($Z = -2.240, p = 0.025$). In addition, the greater ROC-AUC values reflected a more significant change in objective measures. The ROC-AUC values of both RPDE and $D_2$ were greater than 0.7, and were of equal importance for comparison between pre-op and post-op voices.

In present study, two pre-op voice samples could not be used to obtain shimmer, jitter and NHR values due to the extreme aperiodicity. Fig. 2 shows the waveform of these voices and the nonlinear parameters. According to Fig. 2, these voice signals lacked periodicity and subharmonic characteristics and presented with greater chaotic quality.

**TABLE I. SUMMARY STATISTICS FOR ALL MEASURES IN EACH CATEGORY**

<table>
<thead>
<tr>
<th>Measure</th>
<th>Pre-op</th>
<th>One month post-op</th>
<th>Three months post-op</th>
<th>Pre vs. Post 1 Mon.*</th>
<th>Pre vs. Post 3 Mon.*</th>
<th>ROC AUC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>0.915</td>
<td>0.139</td>
<td>0.973</td>
<td>0.846</td>
<td>0.062</td>
<td>0.867</td>
</tr>
<tr>
<td>SD</td>
<td>0.139</td>
<td>0.073</td>
<td>0.973</td>
<td>0.062</td>
<td>0.026</td>
<td>0.128</td>
</tr>
<tr>
<td>Median</td>
<td>0.846</td>
<td>0.062</td>
<td>0.867</td>
<td>0.812</td>
<td>0.047</td>
<td>0.751</td>
</tr>
<tr>
<td>DFA</td>
<td>0.579</td>
<td>0.089</td>
<td>0.549</td>
<td>0.496</td>
<td>0.075</td>
<td>0.439</td>
</tr>
<tr>
<td>RPDE</td>
<td>2.630</td>
<td>0.663</td>
<td>2.347</td>
<td>1.908</td>
<td>0.200</td>
<td>0.013</td>
</tr>
<tr>
<td>Jitter (%)</td>
<td>2.834</td>
<td>3.357</td>
<td>1.543</td>
<td>1.168</td>
<td>0.807</td>
<td>0.025</td>
</tr>
<tr>
<td>Shimmer(dB)</td>
<td>1.042</td>
<td>1.089</td>
<td>1.208</td>
<td>0.488</td>
<td>0.278</td>
<td>0.327</td>
</tr>
<tr>
<td>NHR</td>
<td>0.188</td>
<td>0.285</td>
<td>0.100</td>
<td>0.041</td>
<td>0.052</td>
<td>0.065</td>
</tr>
</tbody>
</table>

*Entries marked (*) are significant at the 95% level.*
IV. DISCUSSION

A number of surgical methods have been documented for managing glottic insufficiency. As a surgical procedure for UVFP, injection laryngoplasty using hyaluronic acid is gaining its popularity among otorhinolaryngologists in Hong Kong. Yet, few studies examined the change in voice quality after the procedure, and the available few only reported outcomes based on English speaking patients. The present study examined the treatment efficacy of injection laryngoplasty among Cantonese iatrogenic UVFP patients. Voice quality was examined by using perturbation analyses and nonlinear dynamical analyses.

Previous studies have demonstrated that injection laryngoplasty improved patients’ voice quality and reduced the mean airflow rates [5, 19, 20]. Umeno et al. found that there was a significant improvement in some speech and voice variables, with voice being restored to preparalytic levels based on acoustical outcomes including harmonics-to-noise ratio, peak prominence, long-term average spectrum, and phonation, pitch perturbation, and amplitude perturbation quotients [20]. However, most of these studies reported surgical efficacy of injection laryngoplasty from non-toral language speaking patients. The only one available study that assessed patients of a tone language was reported by Fang et al. [19]. In the study, fat injection laryngoplasty was found to be effective in enhancing acoustic outcomes and quality of life in UVFP patients of Taiwan. Fang et al. also observed a significant improvement in jitter after one month post-surgery, and the improvement was sustained over a period of 12 months. Harmonics-to-noise ratio values obtained at twelve-month follow-up was significantly superior to that before operation. However, the present study failed to yield any significant improvement in jitter and NHR values, despite the large pre-op and post-op difference in jitter found. Yet, our nonlinear dynamical measures (RPDE & D$_2$) were significantly improved after surgery, confirming the positive surgical outcomes of injection laryngoplasty in patients’ voices.

As a matter of fact, perturbation analyses appeared to be questionable and perturbation measures therefore might not be reliable. The extremely large variability inherent in pathological voice signals rendered perturbation analyses not reliable, even with manual editing by skilled and well-trained researchers [10]. Pre-op UVFP voices were aperiodic and the mathematical framework for linear signal processing simply becomes fallible and inapplicable. In the present study, for example, two of pre-op UVFP voice samples were severely dysphonic, and, as a result of the high aperiodicity of the signals, perturbation analyses could not be carried out. In contrast, nonlinear dynamical analysis was not limited in linear digital signal processing and has shown to be a more reliable and stable analytic method especially for studying aperiodic voice signals [10, 11, 13]. Little et al. investigated the clinical performance of RPDE, D$_2$ and DFA measures for examining changes in voices following therapy [10]. They found that RPDE could illustrate a significant change before and after treatment and were of equal importance. This is consistent with the current findings. RPDE was found to significantly decrease after injection laryngoplasty. In addition, D$_2$ was also significantly reduced after the procedure. The post-operative change in patients’ voices revealed by the nonlinear measures supports the notion that injection laryngoplasty is effective in treating UVFP.

V. CONCLUSION

The present study examined the perturbation and nonlinear dynamical characteristics associated with UVFP patients before and after injection laryngoplasty. Performance of different measures in detecting improvement in voice quality after surgical treatment for UVFP was also quantitatively assessed. It is found that injection laryngoplasty can significantly improve vocal quality in UVFP patients, in a way similar to medialization thyroplasty. Nonlinear dynamical analysis appears to be more superior in evaluating aperiodic signals such as pre-operative UVFP voices. To obtain a more comprehensive depict of injection laryngoplasty, future studies should include more UVFP voice samples, and nonlinear measures, self-perceived quality of life (QOL) and perceptual voice quality should be correlated.

ACKNOWLEDGMENT

Research supported by National Natural Science Foundation of China (NSFC 61135003, NSFC 90920002), Guangdong Innovative Research Team Program (No. 201001D0104648280), and Health and Medical Research Fund (HMRF) (No.09101191).

REFERENCES